17537

NATIONAL LIBRARY
OTTAWA

BIBLIOTHÈQUE NATIONALE
OTTAWA

CANADA

NAME OF AUTHOR........William Bruce Handley

TITLE OF THESIS........Substantive, Structural and External
Components of Construct Validity in
Instrumentation

UNIVERSITY........University of Alberta

DEGREE FOR WHICH THESIS WAS PRESENTED...Ph.D.

YEAR THIS DEGREE GRANTED...1973

Permission is hereby granted to THE NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

(Signed)...W.B. Handley

PERMANENT ADDRESS:

General Delivery
The Pas, Manitoba

DATED...May 7.........19 73

NL-91 (10-68)

THE UNIVERSITY OF ALBERTA

SUBSTANTIVE, STRUCTURAL AND EXTERNAL COMPONENTS OF CONSTRUCT VALIDITY IN INSTRUMENTATION

BY

©

WILLIAM BRUCE HANDLEY

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA
FALL, 1973

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled SUBSTANTIVE, STRUCTURAL AND EXTERNAL COMPONENTS OF CONSTRUCT VALIDITY IN INSTRUMENTATION submitted by William Bruce Handley in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Ram K. Gupta
Supervisor

V. R. Nyberg

John McLeish

B K Sinha

M. P. Strommen
External Examiner

Date May 7, 73

ABSTRACT

This research examined the contribution made by the 'substantive', 'structural' and 'external' components of construct validity (Loevinger, 1957) when employed sequentially and systematically in test development.

The substantive component was applied to develop a scale for each of eight constructs related to youths' 'concerns'. The rationale underlying seven of the constructs is reflected in Mutuality or interpersonal relations amongst youth, the rationale of the eighth pertained to Sense of Mission. Then 113 items were selected by the investigator from the 420 item Youth Research Centre Survey deemed appropriate to the above constructs. The constructs were described in detail, and a panel of 11 judges assigned them to the relevant constructs. The items under each construct constituted substantive valid scales. On 80 of the 113 items, the judges had 82% or higher agreement; on 31 of these, the judges were in complete agreement. Agreement on 26 items ranged from 55 to 73%. Seven items were considered not appropriate for the constructs by the judges.

The structural component of construct validity addressed itself to the question: 'Once a scale has been developed rationally and its content validity ascertained, can it be refined further through examining it psychometrically

with respect to the nature of the intercorrelations of the items assigned to it by the judges? That is, through assessing it in terms of homogeneity.' This was done by cluster analyzing the item variance-covariance matrix of each scale and then by differentially weighting each of the item responses. The responses of 7050 subjects to the items were used for the purpose. Cluster analysis left four scales unchanged; two scales were shortened by one item; one scale by two items; and one scale was subdivided into two separate scales. Differential weighting slightly increased the internal consistency of the scales.

The external component of construct validity was used to ensure that each of the scales possessed high criterion-related validity. Since it was not possible to employ the usual correlational approaches for this purpose, the alternative procedure employed was to make rational predictions and then examine them empirically. The rationale underlying the predictions was that of Mutuality. Criterion groups were formed on the basis of the subjects' responses to ten items of personal data. Since one scale was related to Sense of Mission rather than to Mutuality, the predictions were not expected to be supported in its case, thereby providing indirect evidence about its validity. The predictions were examined in two ways: (1) using the entire scale and (2) using each item within each scale. The predictions were confirmed by all except one scale, the scale which had resulted from the subdivision of a larger scale on cluster analysis. When the items within each scale were examined, all items within four of the scales confirmed all the predictions, the sizes of three scales experienced nominal changes. The items within two scales did not confirm the hypotheses. However, one of the scales was related to a different underlying rationale and as such, indirectly supported the predictions. The scale which was obtained by subdividing a larger scale was the only one which did not support the predictions.

ACKNOWLEDGEMENTS

The author wishes to express his sincere thanks to the members of his committee, Dr. J. Chambers, Dr. R. K. Gupta, Dr. J. McLeish, Dr. V. Nyberg, Dr. B. K. Sinha and to Dr. M. P. Strommen for their assistance in the preparation of this thesis, and to the Youth Research Centre in Minneapolis for the use of the data employed in this investigation. Special thanks are due to Dr. Gupta, the committee chairman, for his encouragement and assistance throughout this study.

Special thanks are also due to my typist and wife, Sandra, who spent many long hours typing the various drafts of this manuscript and to my daughters, Cheryl and Stacey, to whom I wish to dedicate this thesis.

TABLE OF CONTENTS

Chapter Page

I. THE PROBLEM 1

Introduction 1

The Problem 5

Significance of the Study 7

II. REVIEW 9

Introduction 9

Nature of Constructs 10

Constructs in the Context of Validity 13

Empirical Assessment in Structural Validity . . 20

Structural Validity 20

Factor Analysis 22

Assessing Criterion Related Validity 23

III. DESIGN, PROCEDURES AND RESULTS 27

Population of Subjects and Sample 27

Universe of Items 28

The Constructs 29

The Substantive Component of Construct Validity 31

Conclusion Regarding the Substantive Component of Construct Validity 61

The Structural Component of Construct Validity 63

Maximizing the Internal Consistency 71

Summary Regarding the Structural Component of Construct Validity 74

Chapter Page

The External Component of Construct Validity: Using Entire Scales 75

The External Component of Construct Validity: Using Item-Within-Scale Approach . 92

IV. SUMMARY AND CONCLUSIONS 105

Summary 105

Results and Conclusions 108

Limitations of the Study and Suggestions for Further Research 112

REFERENCES 114

APPENDIX A: INSTRUCTIONS FOR RATERS 123

APPENDIX B: TABLES 3.14-3.19 FOR EXTERNAL COMPONENT . 137

APPENDIX C: PROBABILITIES OF F-RATIOS FROM ANALYSIS OF EACH INDIVIDUAL ITEM 144

LIST OF TABLES

Table Page

3.1 SUMMARY OF JUDGES' ASSIGNMENTS OF ITEMS TO CONSTRUCTS 33

3.2 ITEMS ASSIGNED BY JUDGES TO SCALE # 1(a): PEER AND TEACHER RELATIONS 41

3.3 ITEMS ASSIGNED BY JUDGES TO SCALE # 1(b): PEER RELATIONS (OPPOSITE SEX) 44

3.4 ITEMS ASSIGNED BY JUDGES TO SCALE # 1(c): LACK OF SELF CONFIDENCE 47

3.5 ITEMS ASSIGNED BY JUDGES TO SCALE # 1(d): SCHOOL RELATIONS AND PROBLEMS 50

3.6 ITEMS ASSIGNED BY JUDGES TO SCALE # 1(e): PERSONAL WORTH 52

3.7 ITEMS ASSIGNED BY JUDGES TO SCALE # 2(a): FAMILY UNITY 54

3.8 ITEMS ASSIGNED BY JUDGES TO SCALE # 2(b): PARENTAL UNDERSTANDING 56

3.9 ITEMS ASSIGNED BY JUDGES TO SCALE # 3: SOCIAL CONSCIOUSNESS 59

3.10 SCALE RELIABILITIES BEFORE AND AFTER USING THE METHOD OF RECIPROCAL AVERAGES 73

3.11 MEANS AND F-RATIOS FOR SCALE # 1(a): PEER RELATIONS (GENERAL) FOR CONTRASTING GROUPS FORMED ON THE TEN CRITERIA 82

3.12 PROBABILITY MATRIX FOR MULTIPLE COMPARISON OF MEANS FOR HYPOTHESIS NO. 6 FOR SCALE # 1(a): PEER RELATIONS (GENERAL) 85

3.13 MEANS AND F-RATIOS FOR SCALE # 1(b)-1: PEER RELATIONS (OPPOSITE SEX--MARRIAGE) FOR CONTRASTING GROUPS FORMED ON THE TEN CRITERIA 89

Table Page

3.20 MEANS AND F-RATIOS FOR SCALE # 3: SOCIAL CONSCIOUSNESS FOR CONTRASTING GROUPS FORMED ON THE TEN CRITERIA 90

3.21 PROBABILITIES OF F-RATIOS WHEN THE 10 HYPOTHESES ARE TESTED USING EVERY ITEM OF SCALE # 2(a): FAMILY UNITY 95

3.22 PROBABILITIES OF F-RATIOS WHEN THE 10 HYPOTHESES ARE TESTED USING EVERY ITEM OF SCALE # 1(d): SCHOOL PROBLEMS 97

3.24 PROBABILITIES OF F-RATIOS WHEN THE 10 HYPOTHESES ARE TESTED USING EVERY ITEM OF SCALE # 1(b)-1: PEER RELATIONS (OPPOSITE SEX--MARRIAGE) 101

3.29 PROBABILITIES FOR F-RATIOS WHEN THE 10 HYPOTHESES ARE TESTED USING EVERY ITEM OF SCALE # 3: SOCIAL CONSCIOUSNESS 103

LIST OF FIGURES

Figures Page

3.1 Criterion Items Used to Examine the External Component of Construct Validity 77

3.2 Items in Scale # 2(a): Family Unity 94

3.3 Items in Scale # 1(d): School Problems 96

CHAPTER I

THE PROBLEM

Introduction

Data used in research are more meaningful if the instruments used to collect them meet the criteria of good tests. These criteria include the instruments' reliability and validity. Evidence in regard to these do not constitute a serious problem in those situations in which there is a relevant, observable criterion which can be employed as a basis for an examination of the validity. This criterion can be external for the assessment of predictive validity. Alternatively, it can be the content domain, relative to which the developer and the user can judge the adequacy of the content that the instrument is sampling. In many instances, however, there is available neither a pertinent criterion to predict nor a domain of content to sample. Construct validation of the instrument becomes specially crucial or relevant in situations such as these.

The rationale for construct validity of psychological tests was proposed by Cronbach and Meehl (1955) which was probably the first major paper in the area. This rationale is still the widely accepted one. They suggested that construct validation was necessary whenever the trait under consideration could not be directly measured. As a consequence, any assessment must be conducted on the basis of the underlying theory concerning the trait or the construct in question.

Cronbach and Meehl (1955) dealt solely with the theoretical issues related to construct validation and did not suggest specific techniques for its assessment. The non-availability of a summary statistic to represent construct validity and the absence of a specific approach to ascertain it hinder serious use of the concept. As a consequence, the research concerning a test generally constitutes evidence in regard to this type of validity. For example, an instrument that is either newly developed or used for a purpose for which it was not originally developed, must be evaluated for construct validity by obtaining evidence of as many of the other types of validity as possible, specially content and criterion-related validities.

Loevinger (1957) proposed a method of construct validation which employs all the usual types of validation, including content and criterion validation as well as the internal structure, that is, mutual behavior of the items within a test or a scale. She argued that in order to fully assess the construct validity of a psychological test, it was necessary to employ each of the three approaches, sequentially. She called them components. They are: (1) the substantive component, concerned with the formulation of a sound rationale for a construct, operationalizing the rationale by reducing it to behavioral terms or items and then examining them through experts' judgments; (2) the structural component, concerned with the factorial structure or homogeneity of the scale; and (3) the external component, related to appropriately established criterion-related validity.

The test developer should, if he wishes to devise an instrument which possesses construct validity, gather all types of evidence in such a systematic manner. Essentially, the approach is to formulate a sound rationale with respect to the particular concept or construct of interest to him. This necessitates a thorough understanding of the theoretical basis or background of the construct and its incorporation into the testing situation. This rationale also provides a basis for inferring behaviors and then writing or selecting items reflecting these behaviors.

When a pool of such items for the particular construct has been compiled, one should assess their content validity. This requires the involvement of independent judges who are familiar with the underlying rationale and who can competently infer from it the content domain. The result of their decisions or judgments can be expressed in terms of percentage of agreement on each item. On an overall basis, the outcome can be termed the substantive component of construct validity.

The next step after establishing the substantive validity is to examine the scale psychometrically and to refine it further by analyzing the actual responses of the subjects to the items. The usual method is to maximize the homogeneity of the scale by removing items which reduce its 'cohesiveness'. This results in a scale which has high internal consistency, so that the scores have better psychological interpretability.

Once maximum homogeneity or internal consistency has been secured, construct validity is further assessed by determining whether or not the scale possesses criterion-related validity. To do this requires an observable criterion which is usually not available. The alternative is to make logically justifiable predictions using the underlying theory, and then to see whether the predictions are substantiated by data. This serves the purpose of refining the scale further and to improve its applicability.

In practice, the above approach of Loevinger is seldom used when validating tests. For a variety of reasons, most researchers do not proceed systematically in this fashion. Their reasons seem to be related to convenience, shortage of time and money and non-availability of relatively large scale data needed for such a thorough approach. In fact, no research could be traced which may come close enough to this procedure. The question, therefore, arises whether such a thorough approach in test construction is even necessary from the point of view of the expected improvements or contributions. It may be that the work involved in relation to some of the components contributes little to justify it. This study attempts to get some answer to this question.

The Problem

The first step was to formulate a rationale related to certain concerns generally experienced by adolescents: concerns about themselves, their families and the world around them. These three major areas of concern were further subdivided into a total of eight constructs. One hundred and thirteen items which seemed to represent each of these constructs were then selected from a pool of 420 items assembled for a large scale survey by the Youth Research Centre, Minneapolis, Minn., U. S. A. The items reflected youth's perceptions of himself, his family, friends, religion and society. The 113 items were then sorted into eight groups, each related to a given construct by independent judges. Subjective scales were thus formulated on the basis of high agreement among these judges.

The internal consistency of each scale was next examined and maximized by removing those items which did not correlate highly with the other items. The method of Reciprocal Averages was then employed to determine optimal weights for item responses to complete the requirements for the structural component.

Hypotheses were then formulated on the basis of the underlying rationale, using the responses to certain biographical items as criteria. Analyses were completed for each item to determine whether or not these predictions came true.

Briefly, then, the major question to which this research was directed was: In what ways or to what extent does the work related to the structural, substantive and external components of construct validation contribute in the development of an instrument? The major problem reduces to the following minor questions:

1. Can scales be developed on the basis of considerations of substance or content only as examined by knowledgeable judges? It was expected that the scales would be substantiated by high agreement among judges.
2. Do they remain largely the same when subjected to structural examination? It was expected that the latter would lead to some refinement. These refinements could result in improved internal consistency of the scales by removing unrelated items. This would, in turn, improve the psychological interpretability of the scale scores.
3. Do the scales which are sound on the criteria of substantive and structural components have criterion-related validity also? It was expected that these scales would exhibit a high degree of relationship with the criterion.

Significance of the Study

Research in social sciences in field-settings can be described as research generally in community settings for which immediate answers to seemingly pressing problems are required. There is often considerable pressure on the investigator from politicians, public administrators and taxpayers to provide immediate answers to complex problems. Examples of settings of this type are problems relating to economic, cultural and educational poverty and human rights that are the major concerns of agencies such as Alberta NewStart Inc., Indian and Metis Associations, civil rights movements and other social action agencies engaged in what has been termed as the "war on poverty".

Conducting research in such settings is complicated by the fact that the agency in question wants immediate solutions with a minimum of expenditure on experts and expertise. In addition, they have to start with a problem which is usually vast and vague in nature, e.g. 'make the Indian a better person' or 'find ways to get these people back to work'. A typical situation is of the type where the researcher is brought in to assess the effects of some 'super' program. He is usually not involved in planning the project and is requested to simply assess it and is informed that it will commence in a few days. The researcher then must determine what the goal is, develop a rationale and prepare an instrument to measure it. Since he is under pressure of time, short in expertise and money, his usual course of action is to design an instrument on an ad hoc basis, considering only content validity, and that too in a rush, often using himself as the sole judge. This, however, leaves him in the dark in regard to other aspects of construct validity.

It is apparent that a systematic approach to instrumentation which takes into account the substantive, structural and external aspects of construct validity is required. It was the intent of this study to investigate whether such a method of test development does, in fact, bring in sufficient improvement over the results obtained from the usual practice described above.

CHAPTER II

REVIEW

Introduction

During the past 20 years or so, psychologists have been actively refining the research and the techniques of test development and, in the process, paying increasingly more attention to validity. On the side of theory, a great deal of attention has been paid to construct validity. Briefly, construct validity has been defined as "an investigation of the qualities of a test, that is, by determining the degree to which certain explanatory concepts as constructs account for performances on the test". (APA, 1966, pp. 12-13) Construct validity has been of special relevance and interest in the assessment of personality.

Little has, however, been done to apply construct validity in test development. Construct validation of personality tests has been largely related either to comparing correlations with existing scales or to treating it in much the same manner as content validation. The latter is specially true about those who are opposed to construct validity, e.g. Bechtoldt (1959) and Brodbeck (1963). Most psychologists, however, have shown a growing concern over psychological interpretation of tests. This concern began as early as 1951 with Cureton (1951) and Peak (1953).

Nature of Constructs

The increasing concern for theoretical interpretations of test scores necessitated a deeper and clearer treatment of the theoretical underpinnings of the particular attribute which the instrument was attempting to or claiming to measure. This requirement was not necessary when predictive validity was the major consideration because, in the latter, evidence of predictability was all that mattered and neither the content validity of the scales nor the theoretical implications was important. An overstated example might be that the way a subject shone his shoes was a good predictor of engineering success provided a high correlation was found between the two variables. It was not necessary to ascertain a reason for such high correlations--the mere existence of them was sufficient to justify their use.

Early recognition of the usefulness of constructs in testing came partially as a result of the analyses undertaken for identifying the factorial structure or composition of the variables. It was generally agreed that Dimension Analysis (factor and component analyses involving intercorrelations of items) was useful in determining meaningful categories of variables which had no external criterion, that is, were not operationally defined, e.g. ego strength, anxiety, authoritarianism, etc. For variables and notions such as these, interest in the underlying structure of the data forced investigators to devise new methods of analyses.

Peak (1953) proposed that the notion of non-operationally defined terms be defined under the label "Functional Unity". As she postulated, this was similar to the results from factor analyses where the tests are grouped or categorized on the basis of some superficial criteria.

Because there was a general feeling of uncertainty concerning the nature of constructs and construct validity, Cronbach and Meehl, two of the seven members of the APA's original Committee on Psychological Tests, presented a paper which clarified and elaborated on the nature of constructs and its implications in validation of tests. They stated that constructs had three characteristics. First, it was a postulated attribute assumed to be reflected in test performance. Secondly, it had predictive properties. Thirdly, the meaning of a construct was given by the laws with the result that the clarity of knowledge of the construct was a positive function of the completeness of the theory or nomological net.

The notion of a construct was included in the definition of a theory where "a theory consists of constructs and observables, interconnected by statements" (Cronbach, 1971). To illustrate the point, Cronbach (1971) discussed a general theory about expansion of metals. It included the comment that a solid metal expands in length as a linear function of the increase in temperature. The constructs in this example are solid, metal, length and temperature. The model constitutes the statement interconnecting the constructs. For the model to qualify as a theory, the constructs and interconnecting statements must be operationally defined, that is, connected with an indicator or observable data. In this case, copper was tested under various temperatures and a linear expansion coefficient calculated. This was further tested to determine whether or not the coefficient was true under all conditions. This system of interconnected constructs and statements with an external indicator or observable data constitutes the nomological network (Hempel, 1966). This example also illustrates the three characteristics of a construct, viz. (a) it is a postulated attribute, (b) it has predictive qualities and (c) it has meaning (Cronbach and Meehl, 1955).

Torgerson (1958), in his discussion of a well-developed science for social and behavioral sciences, was not different from Hempel's (1966) discussion of theories and constructs. He described a science as comprising of a theory on one hand and empirical evidence (data) on the other. The interplay between the two formulates a viable science.

Margenau (19[illegible]) described a science as having constructs which are related to one another, forming a network of logical relationships. In addition, there are empirical or experimental procedures relating the network of the construct to the observable data. It is the strength of the relationship of the constructs to the data which determines the soundness of the theory.

Cronbach and Meehl (1955) suggested that three sets of laws or relationships constitute a nomological network. They were the interrelations of: (1) observable properties or quantities with each other, (2) theoretical constructs with observables (data) or (3) different theoretical constructs with one another. These laws may be statistical or deterministic (Cronbach and Meehl, 1955, p. 290).

In the social sciences, it is known that the connections or statements between theoretical constructs and observable data are often weak or incomplete. At the same time, theoretical constructs may interrelate highly with each other on some rational basis, but there may be only an indirect connection to the observable data. Cronbach and Meehl (1955) observed that when the nomological network was complete and made connections with observable data, construct validation could occur.

Constructs in the Context of Validity

Although the general notion of construct validity has been around for some time, particularly in personality assessment, it was first given formal recognition as a separate or independent type of validity in 1954 by the American Psychological Association in its "Technical Recommendations". Prior to this, the concept of construct validity was referred to under various labels, including trait validity, factorial validity and statistical validity. The APA's (1954, p. 14) definition was as follows: "Construct validity is normally studied when the tester has no definite criterion measure of the quantity with which he is concerned and must, therefore, use indirect measures. Here the trait and quality underlying the test is of central importance rather than either the test behavior or the score on the criteria."

This rather vague and imprecise definition of construct validity was elaborated upon by Cronbach and Meehl (1955), members of the original APA committee which formulated the recommendations. They described construct validation as being involved whenever a test was to be interpreted as a measure of some attribute which was not otherwise operationally defined. They, however, continued to describe construct validation as the process of assessing the validity of the nomological network of the constructs and connectors within a theory either as a whole or of the propositions from which the inference was derived. This, of course, is not possible unless the nomological network is sound with explicit or 'public' steps of inference, that is, unless the relationship between constructs and observable data is complete or unless the theory has an operational definition. Cronbach and Meehl did not discuss this aspect satisfactorily. In fact, his later writings excluded the term "operational definition" from discussions of construct validity (Cronbach, 1970, 1971).

Campbell offered two kinds of construct validation for consideration. The first of these was consistent with the philosophical approach by Hempel, Peak, Cronbach, Meehl and others which he labelled "nomological validation" (Campbell, 1960, p. 547). His second type of construct validation was equivalent to the older, trait validity which was of interest to the investigators in the area of personality assessment where an apriori defining characteristic was not available. Instead, the investigator had to seek some independent method of getting at the same trait. This independent measure, Campbell stated, was not a true criterion and the validity of both measures was determined by the agreement between the two measures and not by the underlying theory as in the case of "nomological validity".

Loevinger (1957) further extended the scope of construct validity to include content and predictive validities also. To her, construct validity had three components: (1) substantive, (2) structural and (3) external. The substantive component was similar to what other writers referred to as content validity and, in general, appeared to be assessed by rationally examining the validity of the sample content, response specifications or what the subject was required to do and the semantic adequacy of the title of the test (Cronbach, 1971).

Loevinger's second aspect, the structural component, included concepts such as the nature of the relationship between the subjects' responses, functional unity or homogeneity and was based on the premise that persons who scored high on one indicator of a construct ought to score high on another indicator of the same construct. This characteristic is also referred to as convergence of indicators.

Loevinger's third component, the external component, is equivalent to what most investigators call predictive/ concurrent validity and included non-test behavior such as factor patterns, absence of distortions, relation to other tests and intercorrelations with external criteria. The method of assessing this component included the same concepts and procedures as are normally used in studying criterion-related validity. The difference is that with Loevinger's model, the construct and the underlying rationale are of utmost importance in predicting, that is, predictions are based on logic, while in the traditional sense, the correlation of the item with the criterion is of central importance, whether the relationship is logical or not. For example, referring once again to the possibility that the way a person shines his shoes might be a good indicator of engineering success provided the correlation between the two is high.

The absence of references in psychometric literature to Loevinger's position described above implies that either her proposal was not deemed practical or perhaps we failed to recognize its potential usefulness in test development. The

APA's acceptance and incorporation of her ideas into the Standards for Educational and Psychological Tests and Manuals (1966) offers additional justification for this study.

If one were to use Loevinger's prescription for construct validity as a guide in test development, the first step is to insure that the proposed instrument has content validity. The conditions implied are essentially those cited by Cronbach (1971): (1) are the tasks truly representative of the specified universe and (2) was the right universe selected for sampling? Cronbach further asked the following questions about content validity: "Can one explain the test behavior and what implications it has for similar behavior in other situations?" and "Is the universe selected important?" (Cronbach, 1971, p. 452). The first question is related to whether the test follows a carefully developed blue print which is interpretable in an actual situation. The question of importance asks whether or not the item domain was appropriate in the first place.

Melton (1966) cited an example which clearly delineated the difference between construct and content validity. It was in regard to the Watson-Glaser Test of Critical Thinking, where a matter of 'content validity' was to have a qualified person judge whether the test author did indeed assemble a set of problems of the sort they called for in the specifications. On the other hand, to ask a judge whether scores reflected critical thinking was to seek support for his speculation about the construct, "critical thinking".

It is the contention of the present researcher that the substantive component of construct validity lies somewhere between Melton's distinction of construct and content validity. The judges, when assessing the items to be included in the proposed instrument, must rely largely on their judgment of the item in relation to the universe of content, but simultaneously, they must speculate as to the interpretative aspects of the scores based upon the items. This follows logically from the fact that the judges are chosen because they are qualified in a specialized area surrounding the item content. Often, however, they are unable to completely assess the total picture in the context of the nature of the underlying theory and constructs. They, therefore, must pass judgment on the basis of their experience, training or biases as well as their impressions of the nature of the construct.

Another clear example of this aspect comes from Baker and Schulberg (1967). They were dealing with the development and validation of a scale for assessing attitudes toward community mental health. They began by formulating items on the basis of whatever information they could locate concerning community mental health. The resulting pool of 88 items was presented to a panel of judges qualified in the area of attitudes in the context of mental health. They were asked to assess and select those items which were suitable for the proposed scale using rational considerations. Sixty-four items were retained on the basis of high agreement amongst the judges. One could say that those items had more than content validity since the judges' knowledge and experience extended beyond the superficial content aspects and included considerations of clinical or theoretical implications of each item.

There are several methods available to the test developer who wishes to establish the substantive validity of a pool of items. They range from the simplest to the most complex. The simplest is to sort items into groups based upon high agreement between judges. The more complex methods include the Kruskal-Shepard non-metric multi-dimensional scaling technique (1965) and Wiley's Latent Partition Analysis (1969). In these latter methods, the judges must be specially trained in order to understand the complex and time-consuming task entrusted to them. This represents a serious limitation, especially if they are to be used in field settings. A judge can easily render the results meaningless if he does not understand what he is asked to do. The technique also requires a highly sophisticated background in scaling theory and computer facilities for a proper interpretation of the results.

Empirical Assessment of Structural Validity

The structural component of construct validity is concerned with the internal structure of the substantively established scales. This involves an assessment of the internal consistency, homogeneity or functional unity of each scale. The underlying notion is that if a scale has construct validity, it must be reliable. To put it in the converse order, if a scale is reliable, it must measure something--presumably that construct which the judges had agreed upon.

Structural Validity

Structural validity can be represented by any of the following: Loevinger's coefficient of homogeneity, Cronbach's coefficient Alpha, Hoyt's reliability through analysis of variance, and internal consistency through Kuder-Richardson Formula 20 or 21. An algorithm which maximizes the internal consistency while establishing scales is cluster analysis.

The theory underlying cluster analysis or technique of homogeneous keying was given by Loevinger (1947, 1948). DuBois, et al. (1952, 1953) provided a brief overview of the theory as well as a working example. Some examples of its application are provided by Gupta (1968a, 1968b) and Strommen and Gupta (1971).

The technique starts with the variance-covariance matrix of the pool of items which are to be clustered or classified into subgroups or scales. The clusters produced have two properties: (a) each cluster is independent, that is, it has low correlation with any other cluster derived from the same item pool and (b) each cluster has maximum homogeneity or internal consistency. The results thus satisfy Campbell and Fiske's definition of convergent and discriminant validities.

From the variance-covariance matrix, the technique identifies those three items which give maximum reliability, then adds to these three another item which further improves the reliability. This process of adding one item at a time is continued until the addition of an item starts *reducing* the reliability (instead of *increasing* it) by more than an arbitrarily selected small amount, usually .005. The next cluster is formed similarly from the remaining items. Then the third and other subsequent clusters are formed as long as items are available.

The advantages of cluster analysis over factor analysis lie in the fact that it is easy to understand, elegant and straight forward in approach and rigorous in application. Moreover, it can be completed using only a desk calculator, if necessary. As Gupta (1968) and Strommen and Gupta (1971) demonstrated, the results from cluster analysis are very similar to those from factor and component analyses.

Factor Analysis

Without exception, the studies involving construct validation of new instruments which could be located in the literature involved the use of factor analytic techniques. Most of these involved factor analysis of scales. The work of Cattell (1962) and of Guilford (1966) provides the notable examples. In a study of mathematical abilities of grade nine students, Guilford (1966) administered a total of 57 tests to each subject, then factor analyzed these to uncover 14 factors relating to his Structure of Intellect Model. In another study, he factor analyzed 28 tests in order to determine the best predictors of mathematics ability (Guilford, 1965). Cattell (1962) factor analyzed large numbers of tests to determine personality dimensions in children. Hoephner (1967), Hill (1967), Spiegel (1969), Ogston and Drakeford (1971), pointed out the dangers of using factor analysis indiscriminatly on the grounds that many investigators may be misled into believing as if factor analysis uncovered natural dimensions. There is also the shortcoming that the investigator, intentionally or otherwise, can influence his data to behave in a more readily interpretable pattern.

In instrumentation, an additional disadvantage of factor analysis is that advance knowledge is required to use and interpret the results which field researchers rarely have. Also, the reliability of the resulting factors is not immediately known. Moreover, the factors with low internal consistency can be identified. This can mislead the investigator into believing that he has structurally valid scales. It is difficult to develop scales on the basis of content validity as part of the structural component when using factor analysis. Other methods (Cluster Analysis, for example) provide the researcher the option of including or excluding any item.

Any method of instrumentation must be conceptually simple and rigorous. Gupta (1968) and Strommen and Gupta (1971) found that the results from factor and component analyses were very similar to those from cluster analysis. This evidence served as the basis for the decision to use cluster analysis in lieu of factor or component analysis.

Assessing Criterion Related Validity

The concepts of predictive and concurrent validity of the APA (1954) recommendations were incorporated under one heading in the Standards for Educational and Psychological Tests and Manuals (APA, 1966). Many writers, however, still refer to them under two separate headings. The predictive aspect is the most referred to in educational achievement or personnel testing. Over 90 per cent of the papers involving predictive validity of tests fall into either of these two categories. Concurrent validity, on the other hand, is often referred to in relation to tests involving personality or issues in social psychology. Published papers, generally, seemed to comply with Rozeboom's (1965) definition of "diagnostic utility" rather than concurrent validity. Usually the studies are aimed at clinicians who are primarily interested in diagnosis.

Recent studies on predictive validity have tended to relate scores on a test to a follow-up type of criterion. For example, Shevel and Whitney (1969) studied the predictive validity of a new mathematics test which was to be incorporated into the American College Testing battery. Their aim was to establish whether this new instrument would improve the predictive validity of the ACT battery. Such studies rarely show any concern for hypothetical or underlying constructs.

Some studies in the personality area used as the criterion scores on another tried and proven instrument which purported to measure the same trait. For example, Ogston and Drakeford (1971) studied the Costello-Comrey Anxiety Scale (CCAS) by comparing the results with those from the Taylor Manifest Anxiety Scale, the Eysenck Personality Inventory and the Eysenck Neuroticism Scale. In another study, Spiegel (1969) developed a personality inventory and assessed its construct validity by comparing the results with those from instruments such as the California Personality Inventory, Allport and Vernon's Study of Values and the Dogmatism Scale. He found that his instrument was in agreement with the latter so that he could confidently add his particular scale to the already existing scales which purported to measure similar traits.

There are two approaches appropriate for assessing criterion-related validity of a scale--correlational techniques and analysis of variance. The former seems to be common in published research. It is readily applicable in situations where the investigator wishes to assess the validity of his scales by relating the scores from this scale to the scores from another scale. The obtained correlations are in turn judged against an hypothesis relating to expected behavior. A good example is found in McLeish (1970) where he was trying to assess the validity of a series of attitude scales for Teachers' College Students. He chose "formalism" as the concept and constructed an instrument to measure this. Since it was reasonable to expect a high relationship with this to other concepts such as neuroticism, punitiveness, need for financial security and tough-mindedness in education, and since it was also reasonable to expect a low relationship with other concepts such as radicalism, acceptance for change and aesthetic values, these were used as a basis for validating the scale related to "formalism". Thus, validity was sought to be ensured by the confirmation of the apriori hypotheses related to each of the specified scales.

Examples of the other alternative: that of analysis of variance could not be located in the literature. The method seems to have certain virtues in the situations in which the items in a scale are highly related. It can be explained through an example: Suppose one hypothesizes that youth who report to be lonely will score high (high degree of concern) on scales related to peer relationships. To test this hypothesis, one can select two samples--one composed of those who report that they are lonely while the other is composed of those who do not feel lonely. These contrasting groups can be compared using a one-way analysis of variance or 2-sample t-test. A significant difference will substantiate the hypothesis.

The preceding review suggested several methods which could be adopted as part of construct validation of an instrument under construction. For reasons cited in this review, some of the techniques are more readily applicable than others to this investigation and therefore, will be employed. This does not, however, suggest that the other approaches will not give the same results. They will. Since one of the objectives of this study was to offer a simplified method of assessing construct validity, only the most direct and easily interpretable techniques will be considered.

CHAPTER III

DESIGN, PROCEDURES AND RESULTS

Population of Subjects and Sample

Data satisfying the following conditions were needed for this study:

1. Responses to a variety of items representing various constructs in the affective domain.
2. A large number of respondents within a narrow age-range but varying in their backgrounds so that the results could have wider generalizability.

Such data were available from the Youth Research Centre, Minneapolis, Minnesota. They had been collected as part of a nation-wide attitude survey within the United States during 1970. The subjects in the sample constituted more or less a random sample (N = 7050) and ranged in age from 15 to 18 years. They answered an inventory consisting of 420 items. Most of the items related to perceptions and concerns about the world around them.

Every precaution was taken to ensure that all religious affiliations and both sexes from all geographic areas were represented in the sample. This was accomplished through multi-stage random sampling. Even though the survey was designed for youth in the United States, this limitation should not affect the psychometric generalizability of the findings.

Universe of Items

The Youth Research Centre selected or wrote the items for the survey instrument on the basis of the following rationale.

There are two needs which are essential to young people. They are: (1) young people need to feel that they are wanted by others. This need influences what the youth believes about himself, his family, his country and his God, and (2) young people need activities which give them a sense of purpose and significance. It is this sense of mission that gives the young person a feeling of identity. It follows that the person who believes that his life is meaningful and has significance or a sense of purpose will enjoy his work and leisure to a greater extent.

In view of the rationale, the items devised or selected were geared to measure the extent of the young person's perception of himself in two major areas: (a) perception of himself relative to himself and others and (b) perception of himself relative to his purpose in life.

The items are included in two booklets, "Myself and My View of the World" and "My Values and Beliefs". The first of these is concerned mainly with concerns and attitudes that are of interest to teachers and counsellors. It contains 220 items, most of which relate to the subjects' attitudes toward and perceptions of religion, school work, social problems, family, friends and life in general. Some of the items elicit biographical and sociological information also about the subject.

The survey presents the items in the form of declarative statements. For example, 'We are not close as members of a family' or 'I am easily carried away by my emotions'. These reflect the concerns felt by young respondents. The subjects were expected to respond in accordance to a variation of Likert-items. A high score is supposed to reflect high concern.

The second test booklet, "My Values and Beliefs", is designed primarily for use in religious settings. It contains 200 items and is intended to assess the subjects' attitudes concerning religion. These items were not used in the present study.

The Constructs

For the purposes of this investigation, 113 items which were deemed by this researcher as reflecting adolescent concerns were selected. They all came from the first booklet. These constructs relate to 'concerns' which young people have about themselves and the world around them. It seemed reasonable to subdivide these concerns into three major categories which were further subdivisible to yield a total of eight constructs. The major constructs with their subdivisions were as follows:

1. Concerns which an individual has relating to himself.

   (a) Peer Relations (General)

   Concern over relationship with classmates and teachers. Fear of not being accepted.

   (b) Peer Relations (Opposite Sex)

   Concern over relationship with members of the opposite sex including dating and intimate relationships. Fear of not being found desirable or wanted by the members of the opposite sex.

   (c) Lack of Self-Confidence

   Feeling of uncertainty about oneself and fear of making mistakes and appearing ridiculous.

   (d) School Relations and Problems

   Concern over school grades, inability to study or concentrate on school work.

   (e) Personal Worth

   Concern over not having accomplished what he set out to do or having lived up to his ideals. This is characterized by self-criticism.

2. Concerns relating to family.

   (a) Family Unity

   Reflects individual's concern over emotional climate in the home. It includes closeness of family members and consideration for one another.

   (b) Parental Understanding

   Concern over a lack of communication and understanding between the youth and his parents.

3. Concerns relating to social consciousness.

   This category did not have subdivisions. It is related to concerns over what is happening in the country and the world today, especially as it is related to peace and happiness. It includes fear of pollution, nuclear holocaust, war, civil disorder and injustice.

The preceding account is rather brief. A more detailed discussion is given in Appendix A.

The Substantive Component of Construct Validity

The first task in this study was to define the constructs rationally, state them in behavioral terms and operationalize them in terms of items. The constructs were formulated on the basis of existing literature in adolescent psychology in the area of anxiety and concerns.

The 113 selected items were next given to a panel of 11 'qualified judges' who were senior graduate students in the Department of Educational Psychology, University of Alberta, specializing in a wide range of areas, e.g. research design and measurement, counselling, development and learning.

The judges were first familiarized with each of the constructs as defined and then requested to sort each of the items into the eight constructs described to them. The details are given in Appendix A. They were free to sort an item into as many categories as they felt it belonged to. That is, an item could be placed into one, two or more categories.

From Table 3.1, it is noted that the judges exercised this option on 61 of the 113 items. Fifty-seven of the items were placed into two categories by some of the judges, and the remaining four items were placed into three categories by at least one of the 11 judges. The judges were also free to use the 'wastepaper category' in the event they did not feel an item belonged to any of the eight constructs. Table 3.1 shows that this option was exercised with high agreement on three (84, 90, 97) of the 113 items and sporadically on many of the other items. Also, the judges had no time restriction for assigning the items to the constructs. In fact, they were encouraged to sort the items at their own leisurely pace.

TABLE 3.1

SUMMARY OF JUDGES' ASSIGNMENTS OF ITEMS TO CONSTRUCTS

Item	Scale	% agree-ment	Construct Categories 1(a)	1(b)	1(c)	1(d)	1(e)	2(a)	2(b)	3	4	Total assign-ments
1.	2(a)	91%						10	3			13
2.	2(b)	91%						1	10			11
3.	1(c)	73%			8		5					13
4.	1(e)	82%			3		9					12
5.	1(d)	91%				10				1		11
6.	1(a)	64%	7			3	3					13
7.	1(a)	82%	9	1		1	1					12
8.	3	100%								11		11
9.	1(e)	82%					9			2	1	12
10.	3	100%								11		11
11.	2(a)	100%						11				11
12.	2(b)	100%							11			11
13.	1(c)	82%			9		2					11
14.	1(d)	100%				11						11
15.	1(a)	82%	9			2						11
16.	1(a)	100%	11									11
17.	3	100%								11		11
18.	1(c)	55%	2		6		4				1	13
19.	2(a)	91%						10			1	11
20.	2(b)	82%						3	9			12
21.	1(b)	82%	1	9	2						2	14
22.	1(b)	73%		8				1		1	3	13
23.	1(e)	73%	1		2		8			2		13
24.	1(c)	64%	3		7	3		1				14
25.	1(a)	82%	9			1					2	12
26.	1(a)	91%	10		1							11

TABLE 3.1 (continued)

Item	Scale	% agreement	Construct Categories 1(a)	1(b)	1(c)	1(d)	1(e)	2(a)	2(b)	3	4	Total assignments
27.	3	91%	1							10		11
28.	2(a)	100%						11				11
29.	2(b)	82%						3	9			12
30.	1(e)	55%			4		6	2			1	13
31.	1(b)	64%		7	3		2					12
32.	1(c)	100%			11							11
33.	1(a)	55%	6				2		2	1	1	12
34.	1(d)	64%		1		7				4		12
35.	1(c)	100%			11							11
36.	3	100%								11		11
37.	1(e)	82%			2		9				1	12
38.	2(b)	91%	3						10			13
39.	1(b)	82%		9			2				1	12
40.	1(b)	100%		11								11
41.	1(c)	100%			11							11
42.	1(d)	100%				11						11
43.	1(a)	82%	9		1		1				1	12
44.	1(a)	82%	9		1		1				1	12
45.	1(e)	82%			1		9				2	12
46.	1(c)	82%			9						2	11
47.	2(a)	100%						11				11
48.	2(b)	91%		3	1				10			14
49.	1(c)	55%	3		6	1	4					14
50.	1(d)	55%	2		1	6	3					12
51.	1(a)	91%	10		2	1						13
52.	1(c)	55%	3		6						2	11
53.	1(e)	82%			4		9					13

TABLE 3.1 (continued)

Item	Scale	% agreement	Construct Categories 1(a)	1(b)	1(c)	1(d)	1(e)	2(a)	2(b)	3	4	Total assignments
54.	2(a)	91%	1					10				11
55.	2(b)	82%						3	9			12
56.	1(b)	73%		8			1			2	2	13
57.	1(b)	82%	2	9						1	1	13
58.	1(c)	55%			6		4				1	11
59.	1(a)	82%	9		1	1	1					12
60.	1(a)	91%	10								1	11
61.	3	100%								11		11
62.	1(e)	82%			1		9				2	12
63.	3	100%								11		11
64.	2(b)	100%							11			11
65.	2(b)	100%							11			11
66.	1(b)	91%		10	1		1					12
67.	1(e)	55%			1		6				4	11
68.	1(d)	73%			4	8						12
69.	1(d)	82%				9	2				2	13
70.	2(b)	91%					1		10		1	12
71.	--	55%			4		3				5	12
72.	--	55%			3		3			1	5	12
73.	3	100%								11		11
74.	2(b)	100%							11			11
75.	1(b)	91%		10							2	12
76.	2(b)	100%							11			11
77.	1(d)	100%				11						11
78.	1(a)	91%	10		2							12
79.	1(a)	82%	9	1	1		2					13

TABLE 3.1 (continued)

Item	Scale	% agreement	1(a)	1(b)	1(c)	1(d)	1(e)	2(a)	2(b)	3	4	Total assignments
			Construct Categories									
80.	--	50%	3		1		2				5	11
81.	2(a)	100%						11				11
82.	2(b)	100%							11			11
83.	1(b)	100%		11								11
84.	4	55%			3	1	1				6	11
85.	1(d)	82%			2	9	1					12
86.	1(a)	73%	8		2	3						13
87.	1(c)	55%	3		6	2	1					12
88.	3	100%								11		11
89.	1(a)	55%	6		2		2				1	11
90.	4	64%	2				2				7	11
91.	1(a)	82%	9				2				1	12
92.	3	100%								11		11
93.	--	50%	1		1		3	1		2	5	13
94.	1(e)	82%			1		9				2	12
95.	3	100%								11		11
96.	1(e)	82%			3		9					12
97.	4	73%			2		1				8	11
98.	1(e)	55%			3		6				3	12
99.	1(c)	64%	1		7		2			1	1	12
100.	1(c)	73%			8		3	2				13
101.	1(c)	82%			9		3					12
102.	1(e)	73%			1		8				2	11
103.	1(a)-3	64%	7							7	2	16
104.	1(c)	82%			9						2	11
105.	1(b)	82%		9			2			1	1	13
106.	1(e)	73%			2		8				1	11

TABLE 3.1 (continued)

Item	Scale	% agreement	1(a)	1(b)	1(c)	1(d)	1(e)	2(a)	2(b)	3	4	Total assignments
			Construct Categories									
107.	1(b)	82%		9	3			1				13
108.	1(b)	91%		10	3							13
109.	1(b)	91%		10			4					14
110.	1(d)	91%			3	10						13
111.	1(b)	100%		11								11
112.	1(b)	91%		10				1		2		13
113.	3	100%								11		11

Items retained under a construct had to have a minimum level of agreement of 55%, that is, at least 6 out of the 11 judges had to agree that an item belonged to a particular construct. As can be seen from Table 3.1,

(1) This minimum level of agreement was actually used for only 10 or the 113 items.

(2) Only four items had a percentage of agreement less than 55%.

(3) On 29 items, the agreement was 82%.

(4) On 20 other items, the judges had 91% agreement.

(5) The judges were in complete agreement on 31 of the 113 items.

(6) In total, on 80 of the 113 items, the judges had 82% or higher agreement.

(7) It can be further noted that of the 113 items, 18 items were sorted into 1(a), 16 were selected for 1(b), 16 in 1(c), 10 in 1(d), 14 items were placed in 1(e), 7 in 2(a), 13 in 2(b), 13 in 3, seven items were not assigned to any of the constructs, either because of low agreement as was the case for four items or because the judges agreed that the item belonged to the 'wastepaper can category'.

The number of assignments of the item for a particular scale is underlined. Nine judges assigned items to only one category while two judges assigned items to two categories for a total of 13 assignments. Ten of these assignments were placed in Scale # 2(a).

To examine the substantive validity of each item in greater detail, the items were scrutinized in terms of their content relative to the underlying rationale of the scale as well as with respect to the percentage of agreement of the panel of judges who placed the items under the particular scale or construct.

Table 3.2 shows that for 13 of the 18 items of Scale # 1(a), the percentage of agreement was 82% or greater. Of the remaining five items, one had 74% agreement, another 64% and three 55%.

An examination of the contents of the items shows that they relate to the concerns of youth about their peers and adult acquaintances outside the family. Since most of the youth's waking hours are spent in some way associated with the school, it seems reasonable to expect that most of the items concerning peer relations in general would also surround school activities. It is equally important, however, to note that this concern is not directly related to the school itself but to the people associated with school life, including teachers. For example, five of the items relate to teachers (items 1, 3, 10, 11 and 13), illustrating the rather undesirable attitude some teachers seem to have toward adolescents.

The remaining items more directly involve relations with peers, mostly in the school setting. The young person, according to Erikson (1950) and others, feels a strong compulsion to establish close ties with his peers. Those items pertaining to cliques (item 4), classmates (items 2 and 6), recreation (item 5), for example, are designed to assess the degree of concern a young person has concerning the relationship (or lack of it) with others in his age group.

TABLE 3.2

ITEMS ASSIGNED BY JUDGES TO SCALE # 1(a): PEER AND TEACHER RELATIONS

n = 18

Number	Item	% of Agreement
	I am bothered by the fact that....	
1. (6)*	Some teachers act as though a teenager knows practically nothing.	64%
2. (7)*	Classmates at school could be more friendly.	82%
3. (15)*	Some of my teachers are unfair.	82%
4. (16)*	There are cliques (closed groups) in my school.	100%
5. (25)*	We don't always have recreation at school that all can enjoy.	82%
6. (26)*	Some of my classmates are inconsiderate of my feelings.	91%
7. (33)*	Adults tend to underestimate the abilities of young people.	55%
8. (43)*	At school I am often blamed for things I did not do.	82%
9. (44)*	Outside of my family there is no group where I feel I really belong.	82%
10. (51)*	Some teachers are sarcastic and critical of what I do.	91%
11. (59)*	Some of my teachers are not interested in me.	82%
12. (60)*	My interests are often different from those of others my age.	91%
13. (78)*	Some of my teachers do not understand me.	91%

TABLE 3.2 (continued)

Number	Item	% of Agreement
14. (79)*	I do not easily get along with others.	82%
15. (86)*	I feel pressure at school to do what others do.	74%
16. (89)*	I am often jealous of my friends.	55%
17. (91)*	I do not have many friends at school.	82%
18. (103)*	Quite a few in my school are experimenting with drugs.	55%

* Serial number of the item in the questionnaire.

It is noted from Table 3.3 that all the 16 items in Scale # 1(b) have a percentage of agreement of 64% or greater, with 13 of them having 82% or greater.

As the youths reach adolescence, they become increasingly more interested in associations with members of the opposite sex. At the same time, many of them become increasingly concerned with problems related to sex. With age, the concern over marriage and family also increases. This is reflected in those items which relate to seeking a life partner (items 3, 11, 12, 15 and 16). The remaining items, in general, relate to social relations with persons of the opposite sex. These reflect the adolescent's concern over his task of learning appropriate sex roles which involve accepting and learning socially approved male and female roles. According to the literature (Erikson, 1950; Wattenburg, 1956; Garrison, 1965), the greatest problem which adolescents have is that of directing the sex drive into culturally approved channels.

TABLE 3.3

ITEMS ASSIGNED BY JUDGES TO SCALE # 1(b): PEER RELATIONS (OPPOSITE SEX)

n = 16

Number	Item	% of Agreement
	I am bothered by the fact that....	
1. (21)*	I can't seem to think straight when my sexual feelings are aroused.	82%
2. (22)*	I cannot think of sex in marriage as sacred.	73%
3. (31)*	I am afraid of marriage.	64%
4. (39)*	I allow my feelings to overbalance my values in matters of sexual behavior.	82%
5. (40)*	I do not know what a boy (or girl) expects when on a date.	100%
6. (56)*	I find it hard to defend my moral beliefs about sex.	73%
7. (57)*	There are not enough opportunities to be with a mixed group (boys and girls) in social activities.	82%
8. (66)*	I fall in love too easily.	91%
9. (75)*	I cannot stop liking the one with whom I broke up.	91%
10. (83)*	I don't know how boys (or girls) think.	100%
	I wonder about....	
11. (105)*	What to look for in a partner.	82%
12. (107)*	Whether I will marry someone who will give me happiness.	82%
13. (108)*	How to keep boys/girls interested in me.	91%

TABLE 3.3 (continued)

Number	Item	% of Agreement
14.(109)*	Whether my sexual desires are normal.	91%
15.(111)*	Whether I will find a life partner.	100%
16.(112)*	Whether I can find a life partner who feels the same as I do about things that are right and wrong.	91%

* Serial number of the item in the questionnaire.

The scale in Table 3.4 pertains to the concerns about lack of self-confidence. Eleven of the 16 items had a percentage of agreement 64% or more. The items are consistent with Garrison's (1965) and Wattenburg's (1955) notion of lack of self-confidence. It includes fear of embarrassment in public, meeting strange people and, in general, fear of the unknown. An examination of the items also suggests that those on which the judges had only 55% agreement are not centrally relevant to the proposed construct (i.e. items 3, 9, 10, 11 and 12). However, they were included in the scale since they were allotted by the majority of the judges to this scale.

TABLE 3.4

ITEMS ASSIGNED BY JUDGES TO SCALE # 1(c): LACK OF SELF-CONFIDENCE

n = 16

Number	Item	% of Agreement
	I am bothered by the fact that....	
1. (3)*	I am easily carried away by my emotions.	73%
2. (13)*	I worry about little things.	82%
3. (18)*	I am too anxious to please others.	55%
4. (24)*	I lack confidence when reciting in class.	73%
5. (32)*	I am afraid of failure or humiliation.	100%
6. (35)*	I lack the personality and the ability to be a leader of a group.	100%
7. (41)*	I am afraid of making mistakes.	100%
8. (46)*	I do not know what life work to enter.	82%
9. (49)*	I feel that I am not as smart as others my age.	55%
10. (52)*	In a group, I often act differently from what I really am.	55%
11. (58)*	My feelings are easily hurt.	55%
12. (87)*	I lack the ability to participate in sports.	55%
13. (99)*	I do not know what to do when someone makes fun of others.	64%
14. (100)*	I am sometimes so conscious of my faults that I enjoy nothing.	73%

TABLE 3.4 (continued)

Number	Item	% of Agreement
15.(101)*	I am unsure of myself.	82%
16.(104)*	I may become seriously ill or have a crippling accident.	82%

* Serial number of the item in the questionnaire.

Of the ten items listed in Table 3.5 for Scale # 1(d) relating to school and school work, nine had a percentage of agreement 73% or more. Of these nine items, three had perfect agreement.

The rationale of this particular scale is derived from the current economic situation that the young person must continue his schooling until he reaches a certain minimum level, usually senior matriculation. Whereas previously the young person could enter the family farm or his father's small business, the increasing emphasis on technology is forcing most of these small businesses and family farms to discontinue or amalgamate with larger operations. As a result, the adolescent is obliged to go to school if he is to succeed. This, in turn, places additional pressure on him to succeed in school. The prospect of failure and the embarrassment of facing his family and peers afterwards creates a great deal of anxiety in the young person. There is also some resentment concerning irrelevant and uninteresting courses. The young person generally wants to do well in school and is concerned that he is unable to do so because he feels that he cannot study well, concentrate or just does not have the intellectual ability.

TABLE 3.5

ITEMS ASSIGNED BY JUDGES TO SCALE # 1(d): SCHOOL RELATIONS AND PROBLEMS

n = 10

Number	Item	% of Agreement
	I am bothered by the fact that....	
1. (5)*	I have little interest in school studies.	91%
2. (14)*	I do not take my studies seriously enough.	100%
3. (34)*	School requires too much of a person's time.	64%
4. (42)*	I don't know how to study well.	100%
5. (50)*	There are those who are smarter than I am and get better grades.	55%
6. (68)*	I worry about tests.	73%
7. (69)*	School subjects do not offer me enough challenge.	82%
8. (77)*	I have difficulty keeping my mind on my studies.	100%
9. (85)*	I am not satisfied with the grades I get.	82%
	I wonder about....	
10. (110)*	Whether I have the ability to do college work.	73%

* Serial number of the item in the questionnaire.

As noted in Table 3.6, ten of the 14 items chosen by the judges for the scale pertaining to personal worth have a percentage of agreement of 73% or greater.

The intent of this scale is to assess the degree of concern a young person has relative to himself, e.g. fears relating to self-inadequacy, failure in school, etc. These fears and concerns are usually associated with feelings of guilt for not having lived up to the expectations of his family, friends, school and society about him. This feeling reflects itself in most of the items which tend to be of the 'self-reproach' 'self-blame' type.

TABLE 3.6

ITEMS ASSIGNED BY JUDGES TO SCALE # 1(e): PERSONAL WORTH

n = 14

Number	Item	% of Agreement
	I am bothered by the fact that....	
1. (4)*	I become discouraged rather easily.	82%
2. (9)*	I do not feel I help others enough.	82%
3. (23)*	I don't do enough to help others.	73%
4. (30)*	I am not a good example to a younger brother/sister.	55%
5. (37)*	I cannot forgive myself for things I have done.	82%
6. (45)*	I cannot keep from thinking thoughts I feel I shouldn't have.	82%
7. (53)*	I cannot live up to the standards I have set for myself.	82%
8. (62)*	I sometimes have a feeling of superiority.	82%
9. (67)*	Things sometimes do not go the way I want them to.	55%
10. (94)*	I get angry at "little things".	82%
11. (96)*	It seems that I can never do anything right.	82%
12. (98)*	I often feel sorry for myself.	55%
13.(102)*	I day-dream too much.	73%
	I wonder about....	
14.(106)*	Why I behave as I do.	55%

* Serial number of the item in the questionnaire.

The items in Table 3.7 for the scale relating to family unity possess a high percentage of agreement. All items were agreed upon by the judges at a level of 91% or greater. In fact, over one half of the items has unanimity.

In terms of content validity, there is little doubt that the items do, in fact, assess concerns the youth may have about the emotional climate in the home. For example, item # 2 is concerned about the need for a greater feeling of love in the home and item # 6 is concerned about the fact that the mother and the father do not get along as they should.

The importance of this construct is suggested by several adolescent psychologists (Garrison, 1965; Wattenburg, 1955; Erikson, 1950--to name a few) who maintain that even though a young person wishes to be independent of the family, he does want and need to retain emotional ties with the home. This is reflected in a warm emotional atmosphere in the home.

TABLE 3.7

ITEMS ASSIGNED BY JUDGES TO SCALE # 2(a): FAMILY UNITY

n = 7

Number	Item	% of Agreement
	I am bothered by the fact that....	
1. (1)*	We are not close as members of a family.	91%
2. (11)*	We need a greater feeling of love in our family.	100%
3. (19)*	There are not enough social activities in our home.	91%
4. (28)*	My family is not as happy as I wish it were.	100%
5. (47)*	The members of my family are not considerate of each other.	100%
6. (54)*	My father and mother do not get along as they should.	91%
7. (81)*	We do not do things together as a family.	100%

* Serial number of the item in the questionnaire.

Like the scale representing 'family unity', the items in the scale given in Table 3.8 also possessed a high degree of agreement between the judges. All items had a percentage of agreement of 82% or greater.

The content of this scale reflects the adolescent's concern about his parents not understanding him. He has reached the age when he wants to be independent from his parents' restrictions. Erikson (1950) labels this stage of development "ego identity". The parent often does not recognize this and continues to treat him as if he were still a small child. This, of course, only increases the anxiety of the adolescent. As stated earlier, the child is torn in two ways: He wants independence from home and wishes to establish new ties with his peers and attempts to model heroes such as movie stars, racing drivers, teachers and other popular figures. While wanting freedom to spend his own money, to come and go as he pleases and to wear the clothes he likes, the young person also wants to have the protection and security of the home. Scale # 2(a) was devised to assess the latter while Scale # 2(b) is designed to assess the former.

TABLE 3.8

ITEMS ASSIGNED BY JUDGES TO SCALE # 2(b): PARENTAL UNDERSTANDING

n = 13

Number	Item	% of Agreement
	I am bothered by the fact that....	
1. (2)*	My parents seem to have forgotten how it feels to be young.	91%
2. (12)*	It is hard to discuss my problems with my mother.	100%
3. (20)*	My parents (mother or father) nag me.	82%
4. (29)*	My parents (mother or father) try to pry into my private life.	82%
5. (38)*	My parents (mother or father) do not like some of my friends.	91%
6. (48)*	My parents (mother or father) do not understand my dating problems.	91%
7. (55)*	My parents (mother or father) do not let me make my own decisions.	82%
8. (64)*	My father is not as interested in me as I would like.	100%
9. (65)*	My parents (mother or father) force religion on me.	100%
10. (70)*	I do not understand my parents.	91%
11. (74)*	My parents (mother or father) are too strict.	100%
12. (76)*	My mother is not as interested in me as I would like.	100%

TABLE 3.8 (continued)

Number	Item	% of Agreement
13. (82)*	My parents (mother or father) do not trust me.	100%

* Serial number of the item in the questionnaire.

Scale # 3, whose items are given in Table 3.9, is designed to assess social consciousness. A notable feature of this scale was that on all the items, except two (items 4 and 12), the judges had perfect agreement. Item 4 was chosen by 91% of the judges. Item 12, which was also included in Scale # 1(a), had the minimum 55% agreement. Incidentally, this was the only item which the judges had assigned to two categories. When one examines the item, the reason becomes obvious--the item concerns the broader social issue as well as peer relations.

The rationale of this scale is derived from the tendency of young people to be idealistic in their approach to life. They are growing up in a society where they are free to pursue their own beliefs. The mass media and extensive education programs make the youth aware of developments in education, politics and economics. As a consequence, they are inclined to be critical and highly concerned about injustices in their society, pollution, crime and violence, war and other ills of the society.

TABLE 3.9

ITEMS ASSIGNED BY JUDGES TO SCALE # 3: SOCIAL CONSCIOUSNESS

n = 13

Number	Item	% of Agreement
	I am bothered by the fact that....	
1. (8)*	Our national government often seems unresponsive to the needs of the people.	100%
2. (10)*	Pollution of our air and water threatens to destroy all human life.	100%
3. (17)*	Some of the state laws now being enforced are unjust.	100%
4. (27)*	Friends of mine, who don't believe in war, are being forced to enter the military service.	91%
5. (36)*	Our world may be destroyed by a nuclear war.	100%
6. (61)*	Peace among nations seems impossible.	100%
7. (63)*	Revolution and violence may destroy our country soon.	100%
8. (73)*	The ideals of the Constitution are far from the realities of America today.	100%
9. (88)*	Some people want to destroy our government because they find things to criticize.	100%
10. (92)*	There is so much violence and crime today.	100%
11. (95)*	War seems to be a useless butchery.	100%

TABLE 3.9 (continued)

Number	Item	% of Agreement
12.(103)*	Quite a few in my school are experimenting with drugs.	55%
	I wonder about....	
13.(113)*	Whether it is really fair for some men to be unwilling to enter the armed services.	100%

* Serial number of the item in the questionnaire.

Conclusion Regarding the Substantive Component of Construct Validity

It has been seen above that the agreement amongst the judges was high in regard to most of the items and that the instances of lower agreement were limited to three broad categories: self, family and society. The items related to Concerns about the self were the most obvious. For example, some of the judges had difficulty trying to decide whether an item belonged to the category related to self-confidence or to the category related to personal faults. The judges, on the other hand, had no difficulty in deciding whether an item was related to himself or to his family or to his society. The category related to the adolescent's concerns about the society had exceptionally high agreement. An important reason for this could be the absence of subcategories within this group.

The agreement among the judges was too good to believe. Perhaps because the investigator had spent a great deal of time and effort developing the rationale of the various constructs and selecting appropriate items from the Youth Research Centre Survey, there were few items indeed which, on the basis of content, did not seem appropriate for any of the constructs. The use of judges did, however, offer ample support to substantive validity.

The use of a panel of judges rather than depending on the judgment of the investigator alone naturally improved substantive validity. This can be explained in terms of the notion of classification of the sources of variance first discussed by Thorndike (1951) and later expanded by Stanley (1971) which proposes that the sources of unreliability can be partitioned into four components. The first is the lasting-general source which is a function of the underlying trait. Isolating this validates the construct. The three remaining components of sources of variances depend upon the abilities, biases, interests and other extraneous factors possessed by the judges. The use of a larger number of judges helps to reduce the latter and leads to higher reliability and better content validity or substantive component of the scale.

The assignment of the items as done by each member of the panel was also examined. Follow-up discussions with each judge suggested that their assignments were influenced by the factors other than the description of the underlying rationale presented to them in their instructions for assigning items. One judge often departed from the rest of the group and assigned many items to two or three categories. Had he been the sole judge, as is often the case in field settings, the scales would have looked very different. The contribution of a multiplicity of judges is, thus, obvious.

Another possible benefit of using a large number of trained judges arises from the fact that the researcher is obligated to fully develop and clearly explain his rationale.

If the agreement among the judges is very low on a substantial number of the items, then it suggests that perhaps the rationale itself is either not clearly expounded or that the item pool does not represent the 'universe' as defined by the rationale. In either case, the researcher should reconsider his rationale, the constructs, the universe implied by each of the constructs and the particular items that are selected for the scales.

In view of the fact that the judges were in high agreement about the assignment of the items to the various constructs, one could say that the rationale of the constructs used here was clear and that the substantive validity of the scale was high.

The Structural Component of Construct Validity

The next step in developing a set of scales having acceptable construct validity is to consider the major question: 'Once scales have been developed rationally, can they be refined further through examining them psychometrically with respect to the nature of the inter-correlations of the items within each scale?' This amounts to an assessment of the 'structure' of the scales. According to Loevinger (1957), this includes such aspects as homogeneity or functional unity. The latter implies that whereas in the opinion of the judges, the items seem to be alike, do they really 'function' or behave as a team in actual practice.

In this study, the problem was approached through the application of two sequential steps:

1. Correlational analysis of the rationally developed scales to exclude items which detract from the homogeneity or functional unity of the scales.
2. Differential weighting of the responses to the items within each scale in order to maximize the internal consistency of the scales.

For the substantive component, a consensus of the judges had provided the data. For the structural component, the data consisted of the subjects' (N = 7050) responses to the items within each of the scales.

The correlational analysis employed in this study was a conceptually simple and easy to use method of cluster analysis as developed by DuBois, Gleser and Loevinger (1953) and programmed by Gupta and Burnett (1972). The technique selects a nucleus of three items which give maximum possible homogeneity, then adds the fourth item such that the item added improves the homogeneity maximally. Similarly, it continues to add one item at a time until it reaches a point at which the homogeneity of the cluster begins to fall rather than rise with the addition of another item.

The quantitative criterion used was the maximizing ratio given by (1)

$$S = \frac{2\Sigma C_{ij}}{\Sigma V_i + 2\Sigma C_{ij}} \qquad (1)$$

where i = subscript for the items extracted from the item pool, and renumbered 1 to n,

V_i = the variance of item i,

C_{ij} = the covariance of item i with item j,

n = the number of items in a cluster at a given stage of the analysis.

Cronbach's alpha coefficient is a function of the S-ratio since

$$\alpha = \frac{n}{n-1} \frac{2\Sigma C_{ij}}{\Sigma V_i + 2\Sigma C_{ij}}$$

The structural examination of a scale can improve it through one or more of the following ways:

1. A reduction in the size of the scale and, simultaneously, improvement in its reliability.
2. Break down of the scale into two or more parts, one with high and the remaining with high, medium or even poor reliability.

Each of the subjectively and logically derived scales was cluster analyzed, with the following outcomes:

Scale # 1(a)

Item 103 was deleted from Scale # 1(a). It referred to the use of drugs. The other items in the scale are related to a youth's concerns about people around him, (his friends, teachers and other adults) and social relations with people in general. There was already some doubt about the item logically fitting well with the others in the construct and its removal clearly improves psychological interpretability of the scale.

The internal consistency also improved when the item was removed from .84 to .85.

Items 6, 15, 25 and 43 relate to school life and friends. Item 6 refers to youths' perception of the teacher's ability to understand young people; item 15 is concerned with the fairness of teachers; item 25 bemoans the fact that there is not enough recreation in school; item 43 expresses the individual's feeling that he is blamed for things he does not do; item 33 is concerned about adults not appreciating young people. The remaining items reflect concern about peer relations in general, including interests, friends and pressure from peer groups. The total score can be interpreted as concern for relationships with friends, teachers and peers.

Scale # 1(b)

The rationally developed Scale # 1(b) relating to concerns over relationship with the opposite sex broke down into two clusters (called Scale # 1(b)-1 and # 1(b)-2 henceforth), each having sufficiently high reliability and, therefore, deserving to be called a scale. The items related to the more serious concerns of selecting a spouse constituted one scale. It included items related to concerns about what to look for in a life partner, whether or not one would marry someone who will bring happiness, whether or not one would find someone who would share his interests and whether or not one would find anyone at all.

An alpha coefficient of .82 for Scale # 1(b)-1 provides psychometric evidence that the scale represents a factor or construct in its own right.

Scale # 1(b)-2 is related to anxiety and concern about sex and social relations with the opposite sex. It includes items such as 'I am bothered by the fact that I can't seem to think straight when my sexual feelings are aroused' and 'I do not know what a boy/girl expects when on a date'.

An alpha coefficient of .80 attests to the fact that the items are sufficiently homogeneous and, therefore, constitute a separate scale.

The items for the two resulting scales are:

Scale # 1(b)-1

I wonder about....

1. (105) What to look for in a life partner.
2. (107) Whether I will marry someone who will give me happiness.
3. (111) Whether I will find a life partner.
4. (112) Whether I can find a life partner who feels the way I do about things that are right and wrong.

Scale # 1(b)-2

I am bothered by the fact that....

1. (21) I can't seem to think straight when my sexual feelings are aroused.
2. (22) I cannot think of sex in marriage as sacred.
3. (31) I am afraid of marriage.
4. (39) I allow my feelings to over balance my values in matters of sexual behavior.
5. (40) I do not know what a boy (or girl) expects on a date.
6. (56) I find it hard to defend my moral beliefs about sex.
7. (57) There are not enough opportunities to be with a mixed group (boys and girls) in social activities.
8. (66) I fall in love too easily.
9. (75) I cannot stop liking the one with whom I broke up.
10. (83) I don't know how boys (or girls) think.

I wonder about....

11. (108) How to keep boys/girls interested in me.
12. (109) Whether my sexual desires are normal.

Scale # 1(c)

This scale, relating to lack of self-confidence, did not undergo any change as a result of cluster analysis. The psychometric analysis in such a case serves the important purpose of supporting the judges' assignment. Its internal consistency of .80 is evidence of its cohesiveness.

Scale # 1(d)

Two items were dropped from this scale related to school relations and problems. They are items # 34, 'School requires too much of a person's time,' and # 69, 'School subjects do not offer me enough challenge.' This scale is designed to measure the concern of youth for school and the related work. Apparently, the items dropped did not fit the general nature of the scale as a whole. The remaining items are more directly related to personal issues associated with school work, e.g. inability to study, inability to take studying seriously and difficulty in concentrating on school work. This is reflected in items such as 'I have little interest in school studies' (# 5); 'I do not take my studies seriously enough' (# 14); and 'I don't know how to study well' (# 42). The homogeneity of the scale is .83.

Scale # 1(e)

The items for this scale did not change as a result of the cluster analysis of the judges' assignments. As with Scale # 1(c), the psychometric analysis served the important purpose of supporting the judges' assignments. Its internal consistency was .83.

Scale # 2(a)

It related to concerns over family togetherness and harmony within the home. It did not change when cluster analyzed and yielded alpha coefficient .85. It provided ample support to the judges' assignment. It would be recalled that on each item in it, the consensus among the judges was in excess of 90%.

Scale # 2(b)

Like Scales # 1(e), # 1(c) and # 2(a), this scale relating to parental understanding was not altered by cluster analysis. Once again, the coefficient alpha was high (.87).

Scale # 3

This scale pertains to social consciousness and was modified by dropping item # 27, 'Friends of mine who don't believe in war are being forced to enter military service,' as a result of cluster analysis. The internal consistency of the resulting scale is .82.

In summary, the cluster analysis of the judges' scales served two purposes. First, it assisted the investigator in a systematic and objective manner to remove those items on which the responses of the subjects were not highly correlated. This had the overall effect of improving the internal consistency which, in turn, improves the psychological interpretability of the scores on the resulting scales. Secondly, it served to support the judges' decisions on four of the eight scales. These analyses resulted in the deletion of one item from each of Scales # 1(a) and # 3 and two items from Scale # 1(d). Also, Scale # 1(b) was split into two separate scales, each with sufficiently high alpha coefficient.

Maximizing the Internal Consistency

In order to further increase the homogeneity of the scales, it was decided to assign to the responses for the item optimum weights by using the method of Reciprocal Averages (Mosier, 1943, 1946).

Briefly, this method begins with the researcher assigning numerical weights to the item responses as a first judgmental approximation. In this study, the items were assigned weights from one to six for the alternative response items. To obtain the second approximation of weights, the response weights were adjusted in proportion to the mean total score for the particular sub-group of persons giving a specific item response. The total and mean scores for each sub-group on a particular response are once again calculated and weights further adjusted. This procedure continues in successive iterations until no further noticeable change occurs by the adjustments of this type. Summary scores thus obtained yield maximum correlation with each of the items within the scale and, consequently, the reliability of the scores in the sense of internal consistency is maximized.

Each of the refined scales resulting from the cluster analysis was subjected to the method of reciprocal averages and the maximized internal consistency was obtained. The apriori and the final reliabilities are given in Table 3.10.

An examination of Table 3.10 shows that the scales were highly reliable before the responses were weighted optimally. In this case, it was difficult to improve the already present high internal consistencies by the use of the Reciprocal Averages method. Improvement is invariably present, however. It can be recommended, therefore, that where computing facilities are available, it is worth using the method of Reciprocal Averages.

TABLE 3.10

SCALE RELIABILITIES BEFORE AND AFTER USING THE METHOD OF RECIPROCAL AVERAGES

Scale	Reliability before optimum weighting	Reliability after optimum weighting
1(a)	.850	.875
1(b)-1	.816	.839
1(b)-2	.794	.834
1(c)	.832	.862
1(d)	.794	.846
1(e)	.826	.844
2(a)	.847	.862
2(b)	.872	.899
3	.817	.864

Summary Regarding the Structural Component of Construct Validity

It is noted that the structural criteria of construct validity, when applied to the judges' scales, modified as many as four of the eight scales. In one instance, Scale # 1(b): Peer Relations (Opposite Sex), the scale was even subdivided into two separate scales, each having high alpha. These two scales respectively consider concerns related to marriage and those related to casual dating and sex. In marriage-related concerns, the young people appear to be afraid of making a poor choice of a partner. They are concerned about the compatability with their spouse. On the other hand, in social and casual relationships with the opposite sex, young people appear to be concerned with such things as controlling their sexual feelings and drives, impressing members of the opposite sex and participating in social activities with others.

The three other scales which were slightly modified as a result of the cluster analysis were Scale # 1(a): Peer Relations (General); Scale # 1(d): School Relations and Problems and Scale # 3: Social Consciousness. These were altered through the deletion of one or two items as a result of which, internal consistency showed some improvement. Scale # 1(a) was reduced by one item, with logical justification as well. Scale # 1(d) shortened by two items, again with sufficient rational justification.

While many of the changes in the scale via the structural component criteria seem minimal, they are nevertheless essential in having a closer look at the substantive validity of the scales. It starts becoming clearer why many of the excluded items are rejected when one re-examines the items in light of this new information. Scales which had no changes are further verified and one can then proceed with his research with greater assurance that he is dealing with well defined constructs.

External Component of Construct Validity:
Using Entire Scales

The confirmation of construct validity requires information from as many sources as possible, including what Loevinger called the external component. It comes close to "criterion-related validity" as described by the APA (1966). The latter can be assessed either by using a predictive criterion or a concurrent one. This research will use the concurrent criterion.

The method of concurrent validation is the one most frequently employed by psychologists and educators when assessing the construct validity of a test. It is often accomplished by correlating the scores on an accepted, proven, well-known instrument designed to measure the same construct to the scores on the new instrument which, supposedly, has certain practical or administrative advantages.

In the absence of such instruments for measuring the constructs of interest in this research, it was not possible to examine concurrent validity through the correlational approach. Therefore, an alternative method, and probably a more justifiable one, was adopted here. In a nutshell, it amounted to making rationally and theoretically justifiable predictions and then examining them empirically to see whether they are substantiated or not. For this purpose, behaviors shown independently of the scales were used as criteria. These were elicited on selected items of biographical and sociological nature concerning the subject. They are listed in Figure 3.1. The responses on these items were obtained at the same time at which the items which formed the scales were administered. In this respect, it is concurrent validation.

The approach adopted here is direct and also ingrained in the theory underlying the constructs. As such, it seems preferable to the usual correlational approach since the latter could easily mean that one is simply duplicating the accepted 'mistakes' of the previous researchers. For example, if one constructs a new instrument designed to measure intelligence and correlates the scores with those from some well known instrument, such as the Stanford-Binet Intelligence Scale, then a high correlation could be merely a measure of success in replicating the earlier but outmoded concept of intelligence used by Terman and Binet.

Number	Item	Classifications
1. (13)*	I tend to be a lonely person.	1. yes 2. no, sometimes
2. (20)*	I find life exciting and full of fun.	1. no, sometimes 2. yes
3. (95)*	My parents are too strict.	1. very much 2. somewhat 3. never
4.(157)*	I have considered suicide.	1. yes 2. no
5.(181)*	How many close friends do you have, people whom you feel really care about you?	1. fewer than 3 2. 3 to 7 3. 8 or more
6.(192)*	To what degree do you feel trusted by your parents (or guardian)?	1. not at all 2. somewhat 3. very much
7.(199)*	Choose the number that best illustrates where you are in your circle of closest friends.	1. periphery 2. in between 3. centre
8.(209)*	I have trouble getting along with my father.	1. yes 2. no
9.(210)*	I have trouble getting along with my mother.	1. yes 2. no
10.(220)*	We have had serious difficulties in our home (prolonged illness, unemployment, death or injuries, personal problems) during the past year.	1. yes 2. no

* Serial number of item in the Youth Research Centre Survey

Figure 3.1. Criterion Items Used to Examine the External Component of Construct Validity.

The examination of the external component of construct validity consisted of two aspects. First, the predictions were examined using scores on an entire scale. Secondly, they were examined, using the responses to each item within that scale based upon the logic that what a scale does as a whole should also be done by each of its parts, in this case, each item. It is implied here that the validity of a scale can be improved by deleting those items, if any, which fail in this respect. The resulting shorter scale would have higher validity, though it could have marginally lower reliability.

The rationale used for this purpose was as follows: Youth who lack a sense of mutuality with others will obtain scores which are significantly different on each of these scales (and on each item within each scale) than those who experience this mutuality. That is, it is expected that youth who feel that they do not belong to some group such as family, circle of friends, teachers, will experience a certain level of concern. At the same time, young people who are fully accepted and trusted by their peers, parents, teachers and friends, both adult and youth, will be significantly less concerned about themselves, their families or the current state of their society and country.

More specifically, the following research hypotheses were set up.

1. Youth who report that they tend to be lonely (item # 13) are more likely to score higher on the scales than those who do not so report.
2. Youth who report that they do not find life exciting (item # 20) will score higher than those who report that they find life exciting.
3. Youth who feel that their parents are too strict (item # 95) will score higher on the scales than those who feel that their parents are not strict.
4. Youth who report that they have considered suicide (item # 157) will have higher scores than those who have never considered such.
5. Youth who feel that they have few close friends (item # 181) will obtain higher scores than those who feel otherwise.
6. Youth who do not feel trusted by their parents (item # 192) will obtain higher scores than those who feel that their parents do trust them.
7. Youth who feel that they are on the periphery of their circle of friends (item # 199) will obtain higher mean scores on the scales than those who feel that they are near the center of this circle of friends.
8. Youth who feel that they have trouble getting along with their fathers (item # 209) will obtain higher scores than those who do not feel this way.

9. Youth who report that they have trouble getting along with their mothers (item # 210) will obtain higher scores than those who do not feel this way.
10. Youth who come from homes which have had serious difficulties in their homes (item # 220) during the past year will score higher on scales for assessing concerns than those who did not have such difficulties.

It will be recalled that there were nine scales in all when the structural component of construct validity was examined. In terms of the two basic premises of the Youth Research Centre Survey--Mutuality and Sense of Mission--the first eight of these scales belong to the former and the last one, Scale # 3: Social Consciousness, to the latter. The ten predictions described above are based upon the subject's feelings about himself as a person and, as such, would probably not be substantiated by Scale # 3. That scale was also examined, however, to see whether it supports the research hypotheses or not.

The ten criterion items listed in Figure 3.1 correspond to the above research hypotheses. Contrasting groups were formed on the basis of the subjects' responses to each of the ten criterion items and then were employed to test the null version of the research hypotheses.

Ten sets of contrasting groups were formed by sorting the responses to each of the ten criterion items. In most cases (items # 13, 20, 157, 209, 210 and 220), the set consisted of two groups. In some (items # 95, 181, 192 and 199), it has three groups. For example, in the two group sets, all those who responded 'no' or 'sometimes' were sorted into one group, while those who responded 'yes' were sorted into the other contrasting group. These included items # 13, 20, 157, 209, 210 and 220.

The scores of the subjects within each contrasting group were analyzed by one-way analysis of variance. They provided the necessary evidence to enable the researcher to either reject or not reject a null hypothesis. A significant F-ratio implied that the corresponding research hypothesis had been confirmed or sustained.

Let us consider Scale # 1(a) in detail. It relates to the concerns about relations with peers and teachers. The underlying rationale of the ten hypotheses assumes that the youth who are anxious about their relationship with others will manifest the anxiety in high scores on the scale. On the contrary, youth who experience warmth, love and acceptance by their friends, both peer and adult, will be less anxious and hence will obtain lower scores on this scale.

The results are given in Table 3.11. It shows that each of the ten research hypotheses is clearly substained.

Criterion item # 13 (Fig. 3.1) is the basis for hypothesis no. 1, 'Youth who report that they tend to be lonely will score higher than those who report that they do not

TABLE 3.11

MEANS AND F-RATIOS FOR SCALE # 1(a): PEER RELATIONS (GENERAL) FOR CONTRASTING GROUPS FORMED ON THE TEN CRITERIA

Hypothesis	Means			F-ratio	Probability
	1	2	3		
1. (13)*	34.49	30.40		255.52	.00
2. (20)*	33.08	29.44		338.86	.00
3. (95)*	32.57	30.42	27.26	168.86	.00
4. (157)*	37.15	33.44		301.02	.00
5. (181)*	33.31	31.49	29.42	101.91	.00
6. (192)*	32.05	32.96	30.17	82.17	.00
7. (199)*	34.08	31.75	30.04	84.73	.00
8. (209)*	36.60	33.86		146.30	.00
9. (210)*	36.39	34.00		106.42	.00
10. (220)*	35.69	34.15		47.50	.00

* Serial number of item in the Youth Research Centre Survey

tend to be lonely'. The prediction is borne out as can be seen from Table 3.11. Those who tend to be lonely have a mean score of 34.49, while those who report that they do not tend to be lonely obtained a significantly lower mean score of 30.40. These results comply with the rationale that persons who tend to be lonely are concerned about being accepted by others.

The same rationale applies to hypothesis no. 2, 'Youth who find life full of fun and exciting will not be as anxious and concerned about relations with peers and teachers and, therefore, will not obtain scores as high on Scale # 1(a) as those who do not find life exciting. This is confirmed by the significantly different mean scores of 33.08 and 29.44.

Persons who feel that their parents are too strict will manifest, in general, a higher degree of concern over relationships with others (friends and teachers included) than those who feel that their parents are not too strict. This is the basis for hypothesis no. 3 on Scale # 1(a). It is noted that this hypothesis employs three contrasting groups. As indicated in Figure 3.1 for criterion item # 95, the classifications, in general, include (1) quite a bit, (2) somewhat, (3) never. The results in Table 3.11 show that the null hypothesis is clearly rejected, thus confirming the research hypothesis that 'Youth who feel that their parents are too strict will score higher than those who feel that their parents are not strict'.

Hypothesis no. 4, 'Youth who report that they have considered suicide will obtain higher mean scores than those who have never considered it', is formulated on the premise that depressed people have a high level of concern over relationships with friends and teachers. Scale # 1(a) is tested on this hypothesis by comparing the subjects' scores from the contrasting groups comprising those who have considered suicide versus those who have not considered it. The results in Table 3.11 confirm the research hypothesis as stated above.

Presumably, those young people who have few friends will be more concerned about their relations with others than those who have many friends. This hypothesis (no. 5) was tested, using the three contrasting groups based on criterion item # 181 (Fig. 3.1). These contrasting groups include those subjects with (i) 8 or more friends, (ii) 3 to 7 friends and (iii) fewer than 3 friends. The results from Table 3.11 for Scale # 1(a) confirmed it.

It was expected that the degree to which a subject felt trusted by his parents would be related to the score he would obtain on Scale # 1(a). Criterion item # 192 served as the basis for this prediction. Three contrasting groups were employed in the analysis including (1) not trusted at all, (2) somewhat trusted and (3) very much trusted. Examination of Table 3.11 for hypothesis no. 6 shows that there was a significant difference between the three categories on the mean scores on Scale # 1(a). However, the results in one case somewhat departed from the prediction.

Comparisons of all possible pairs of the three contrasting groups for hypothesis no. 6 on Scale # 1(a) are given in Table 3.12. It is noted that the means of group 1 (not at all) and group 2 (somewhat) are not significantly different. At the same time, it is shown that the differences between the means of group 1 (not at all) and group 3 (very much) are significant. The difference between the means of group 2 (somewhat) and group 3 (very much) is also significant.

TABLE 3.12

PROBABILITY MATRIX FOR MULTIPLE COMPARISON OF MEANS FOR HYPOTHESIS NO. 6 FOR SCALE # 1(a): PEER RELATIONS (GENERAL)

Contrast Groups	Contrast Groups 1	2	3
1	1.00	.06	.00*
2	.06	1.00	.00*
3	.00*	.00*	1.00

* Significant beyond .01 level.

From these results it is obvious that research hypothesis no. 6 has been confirmed on the basis of two categories--those who feel trusted by their parents versus those who do not feel trusted.

Criterion item # 199 relates to the subject's perception of where he feels that he fits in his circle of friends. As noted from Figure 3.1, responses to this item were used to generate contrasting groups based on where in his circle of friends he feels he fits. These are (1) on the periphery, (2) in between the centre and the periphery and (3) in the centre. Hypothesis no. 7 postulates that those who feel that they are on the outside of their circle of friends will obtain high scores on Scale # 1(a) which is designed to assess their level of concern over relations with their friends. This is based on the assumption that young people need and seek close companionship with others. In this particular case, the research hypothesis that 'Youth who feel that they are on the periphery of their circle of friends will score higher on Scale # 1(a) than those who feel that they are at the centre of their circle of friends' is confirmed as indicated in Table 3.11 for hypothesis no. 7.

Hypotheses no. 8 and no. 9 relate to the subjects' relationships with their parents, i.e. 'Youth who have trouble getting along with their fathers (criterion item # 209) or their mothers (criterion item # 210) will obtain higher scores than those who do not have trouble getting along with their parents'. The premise of mutuality suggests that youth who experience a relationship of warmth and love with their parents would be less concerned about relationships with others, including friends, than those who do not experience this type of relationship.

Examination of the results for Scale # 1(a) in Table 3.11 on hypotheses no. 8 and no. 9 clearly substantiates the two research hypotheses.

Serious difficulties at home, such as prolonged illness, unemployment, death, injuries or personal problems, can be very upsetting to an adolescent. This affects his relationship with his family which is also reflected by a feeling of uncertainty in his relationship with others around him. Thus, it was reasonable to expect that adolescents who have experienced such difficulties will obtain higher scores on Scale # 1(a).

This, in fact, turned out to be true as the F-ratio for hypothesis no. 10 in Table 3.11 shows. Those who reported having had difficulties in their home during the past year obtained a significantly higher score than those who reported not experiencing any serious problems in their home.

The remaining eight scales were examined similarly. The results related to six of them--Scales # 1(b)-2, # 1(c), # 1(d), # 1(e), # 2(a) and # 2(b)--were identical to those for Scale # 1(a).

Each of the 10 research hypotheses was confirmed for each of these six scales. Since the results seem to be repetitive in nature, they are given in Appendix B. The nature of the results for Scales # 1(b)-1 and # 3 was different, however, to a substantial extent, as shown below.

Table 3.13 shows that for Scale # 1(b)-1, hypotheses nos. 5, 6 and 7 were not rejected. Hence, it cannot be implied that there is a relationship between the scores from this scale which relate to concerns about marriage and the subject's perception of where he fits in his circle of friends or the number of close friends he has. Also, whether or not a young person is trusted by his parents is not related to his score on the scale related to marriage.

The results given in Table 3.20 (Tables 3.14 to 3.19 constitute Appendix B) show that the comments given above for Scale # 1(b)-1 also apply to Scale # 3. That is, neither the subject's perception of where he fits in his circle of friends nor the number of friends he has has a relationship to his concern for society. As in Scale # 1(b)-1, there was no relationship between degree of parents' trust and scores on the scale related to social consciousness.

In summary, the rationale upon which the scales were developed and upon which the research hypotheses were formulated is as follows:

Adolescents become more independent as they grow older and seek the warmth and companionship of others outside of

TABLE 3.13

MEANS AND F-RATIOS FOR SCALE # 1(b)-1: PEER RELATIONS (OPPOSITE SEX--MARRIAGE) FOR CONTRASTING GROUPS FORMED ON THE TEN CRITERIA

Hypothesis	Means 1	Means 2	Means 3	F-ratio	Probability
1. (13)*	41.50	39.41		30.11	.00
2. (20)*	40.43	39.23		16.61	.00
3. (95)*	41.59	38.81	35.01	121.22	.00
4. (157)*	41.66	38.99		71.37	.00
5. (181)*	39.85	40.16	39.34	2.91**	.17
6. (192)*	39.45	40.05	39.65	1.84**	.35
7. (199)*	40.03	39.89	39.74	0.23**	.46
8. (209)*	41.29	39.26		37.37	.00
9. (210)*	40.85	39.46		16.50	.00
10. (220)*	40.97	39.33		25.64	.00

* Serial number of item in the Youth Research Centre Survey

** Not significant

TABLE 3.20

MEANS AND F-RATIOS FOR SCALE # 3: SOCIAL CONSCIOUSNESS FOR CONTRASTING GROUPS FORMED ON THE TEN CRITERIA

Hypothesis	Means 1	Means 2	Means 3	F-ratio	Probability
1. (13)*	39.77	38.92		9.23	.00
2. (20)*	39.54	38.66		16.00	.00
3. (95)*	40.60	35.55	29.55	1684.44	.00
4. (157)*	36.58	34.13		83.03	.00
5. (181)*	38.83	39.37	38.94	2.35**	.27
6. (192)*	38.77	39.33	39.02	1.25**	.41
7. (199)*	38.51	39.01	39.26	2.19**	.17
8. (209)*	36.47	31.32		58.95	.00
9. (210)*	35.94	34.55		23.31	.00
10. (220)*	36.07	34.40		36.75	.00

* Serial number of item in the Youth Research Centre Survey

** Not significant

their homes and families (including other adults such as teachers as well as peers). During this period, they begin to develop strong relations with the opposite sex. In addition, they are afraid of humiliation, particularly in public, and do not wish to make themselves look foolish or awkward. Related to this is the fear of not having met the standards and goals set for them either by themselves or by others such as parents or society at large. They also eagerly want to do well in school and are concerned at the prospect of failure to achieve this.

The adolescent changes from a small child who is completely dependent on his parents for his every emotional and physical need to one who eagerly wants to be free, to earn and spend his own money and to seek his own friends. The reluctance of the parents in allowing freedom to their children often results in conflicts between the youth and his parents, thus creating a considerable amount of anxiety for both the child and the parent.

While the adolescent is seeking to break away from home and be free, he still desperately needs the love, understanding and warmth of his parents. Thus, he is torn between wanting to be free and wanting to remain in the comfort of his home.

From this rather brief account of the rationale underlying the constructs and the predictions made on this basis, the concurrent validity of Scales # 1(a), 1(b)-2, 1(c), 1(d), 1(e), 2(a) and 2(b) has been established beyond a reasonable doubt. Some doubt is left about Scales 1(b)-1 and 3, however.

Scale # 1(b)-1 relates to concerns about marriage. It would be recalled that originally it was a part of the judges' Scale # 1(b): relationships with the opposite sex. The fact that it did not support three of the ten predictions suggests some weakness in it. The same applies to Scale # 3. It could, of course, be the weakness in the rationale, too, but more about it in the next section.

In general, it can be concluded that the use of research hypotheses based on rational considerations was successful in establishing the external component of construct validity of the scales. This approach seems to make some contribution to test development and can, therefore, be recommended to researchers.

External Component of Construct Validity: Using Item-Within-Scale Approach

The second step in assessing the external component of construct validity was to examine the same ten rational predictions, using each individual item within each scale. It is assumed that what the item does as a whole is also done by each of its parts, that is, the items.

Procedurally, the responses to each item in a scale were analyzed in exactly the same manner as were the scores on the entire scale. Also, the same contrasting groups and the same research hypotheses were used for this purpose.

To facilitate the presentation and discussion of the results, two examples are selected. These are Scales # 1(d) and # 2(a). Scale # 2(a) (Fig. 3.2) was selected because it represented those scales which gave significant F's for all the ten null hypotheses. The items in the scale behaved as expected, that is, each of the items supported each of the ten research hypotheses, as Table 3.21 shows. By implication, it supported the underlying rationale of the constructs, too.

Scale # 1(d) (Fig. 3.3) was chosen because it represented those scales on which each of the research hypotheses was confirmed, although, not all of its items, when examined similarly, did so. It was the intent of this investigation to identify such items.

The results of testing the null versions of the ten research hypotheses, using each item within Scale # 1(d), are presented in Table 3.22. The F-ratios (from analyses of variance) which are non-significant at $\alpha = .05$ are underlined for convenience. These are interpreted as failing to confirm the research hypotheses.

All the items for Scale # 1(d), and the order in which they were included in the scale at the time of cluster analysis, are presented in Figure 3.3.

I am bothered by the fact that....

1. (1)* We are not close as members of a family.
2. (1)* We need a greater feeling of love in our family.
3. (5)* There are not enough social activities in our home.
4. (1)* My family is not as happy as I wish it were.
5. (2)* The members of my family are not considerate of each other.
6. (4)* My father and mother do not get along as they should.
7. (3)* We do not do things together as a family.

---

* Order in which the items were added to the scale on cluster analysis.

Figure 3.2. Items in Scale # 2(a): Family Unity

TABLE 3.21

PROBABILITIES FOR F-RATIOS WHEN THE 10 HYPOTHESES ARE TESTED USING EVERY ITEM OF SCALE # 2(a): FAMILY UNITY

Item of	Criterion Items									
Scale # 2(a)	13	20	95	157	181	192	199	209	210	220
1. (1)*	00	00	00	00	00	00	00	00	00	00
2. (11)*	00	00	00	00	00	00	00	00	00	00
3. (19)*	00	00	00	00	00	00	00	00	00	00
4. (28)*	00	00	00	00	00	00	00	00	00	00
5. (47)*	00	00	00	00	00	00	00	00	00	00
6. (54)*	00	00	00	00	00	00	00	00	00	00
7. (81)*	00	00	00	00	00	00	00	00	00	00

Note: Decimal is understood before entry.

* Serial number of the item in the questionnaire.

I am bothered by the fact that....

1. (3)* I have little interest in school studies.
2. (1)* I do not take my studies seriously enough.
3. (1)* I don't know how to study well.
4. (4)* There are those who are smarter than I am and get better grades.
5. (5)* I worry about tests.
6. (1)* I have difficulty keeping my mind on my studies.
7. (2)* I am not satisfied with the grades I get.

I wonder about....

8. (6)* Whether I have the ability to do college work.

---

* Order in which the items were added to the scale on cluster analysis.

Figure 3.3. Items in Scale # 1(d): School Problems

TABLE 3.22

PROBABILITIES FOR F-RATIOS WHEN THE 10 HYPOTHESES ARE TESTED USING EVERY ITEM OF SCALE # 1(d): SCHOOL PROBLEMS

Item of Scale # 1(d)	Criterion Items 13	20	95	157	181	192	199	209	210	220
1. (5)*	00	00	00	00	00	00	00	00	00	00
2. (14)*	00	00	00	00	12	00	04	00	00	00
3. (42)*	00	00	00	00	00	00	00	00	00	00
4. (50)*	00	00	00	00	00	00	00	00	00	00
5. (68)*	87	00	00	10	72	27	06	19	99	00
6. (77)*	00	00	00	00	03	00	04	00	00	00
7. (85)*	00	00	00	00	00	08	00	00	00	00
8. (110)*	00	00	00	00	29	00	00	00	12	22

Note: Decimal is understood before each entry.

Underlined entries indicate non-significant F-ratios when $\alpha = .05$.

* Serial number of the item in the questionnaire.

From Table 3.22, it is noted that the F-ratios for item # 1 were clearly significant for all the null hypotheses. It can be inferred that item # 1, 'I have little interest in school studies', fully supports the underlying rationale of the construct for the scale. The same applies to items 3, 4 and 6.

Item # 2, 'I do not take my studies seriously enough', gave significant F's for nine of the ten null hypotheses, the exception being that arising from # 181 (number of close friends). As a consequence, one can fairly confidently retain this item in the scale. For the same reason, item # 4 can also be retained.

The F-ratios for item # 5, 'I worry about tests', were significant for only three of the ten hypotheses, and on as many as seven, they were non-significant. This raises very serious doubt as to the appropriateness of this item for the scale. The recommendation, therefore, is to delete it from the scale. In such cases, one could possibly argue or raise the question: Is it the item which is at fault or is the rationale itself deficient? The contention of this researcher is that it is the former case and not the latter. The reason is that if the rationale itself were faulty, then for one or more hypotheses (or predictions), we would have obtained non-significant F-ratios for several items (underlines would have been along columns in Table 3.22) rather than such F's being for items and, therefore, along rows.

Item # 8, 'I wonder about whether or not I have the ability to do college work', requires careful consideration as to whether one should retain it in the scale or not. It gave non-significant F's for three of the ten null hypotheses. One would probably hesitate to retain such an item within the scale.

From Figure 3.3 it is noted that the items (# 5 and # 8) about which doubt has been raised above, were the last two items to be included in the scale on cluster analysis. It could have been the stage at which "functional drift" had probably started (DuBois, et al., 1952).

Results from item-wise analyses for each item from the remaining scales are presented in Tables 3.23 and 3.25-3.28 of Appendix C. The F-ratios for the items within each of Scales # 1(a): Peer Relations (General); # 1(c): Self-Confidence and # 2(b): Parental Understanding were generally significant for each of the ten null hypotheses. In this sense, they were similar to the results from Scale # 2(a). The non-significant results were few and did not seem to have a pattern or system.

A few items in Scales # 1(b)-2: Peer Relations (Opposite Sex--Social) and # 1(e): Personal Worth behaved in a manner similar to items 5 and 8 or Scale # 1(d) described above. Three of the ten null hypotheses were not rejected on item # 40, 'I don't know what a boy (or girl) expects when on a date', of Scale # 1(b)-2. The same applied to four of the ten hypotheses in regard to item # 75, 'I cannot stop liking the one with whom I broke up', of the same scale.

Item # 30, 'I am not a good example for my younger brother/sister', and # 62, 'I sometimes have a feeling of superiority', of Scale # 1(e) behaved the same way as items # 40 and # 75 respectively of Scale # 1(d). These four items could be considered as being not very appropriate for the scales and the related constructs.

It will be recalled that on Scale # 1(b)-1: Peer Relations (Opposite Sex--Marriage) and # 3: Social Consciousness, three of the ten null hypotheses were not rejected (Tables 3.13 and 3.20). It was expected, therefore, that a substantial number of their items would behave in essentially the same way, that is, yield non-significant F-ratios. This actually happened on several items.

The results of testing the null versions of the research hypotheses, using items within Scale # 1(b)-1 are presented in Table 3.24. The F-ratios which are non-significant are underlined. Examination of Table 3.24 by rows shows that the F's for two (# 2 and # 4) of the four items in the scale are non-significant on three or more criterion items. Columnwise, the F's for criterion items # 192, 'degree of trust by parents', and # 199, 'circle of friends', are non-significant on three of the four items. The same is true on two of the four scale items for # 157, 'I have considered suicide', and # 181, 'number of close friends'.

TABLE 3.24

PROBABILITIES FOR F-RATIOS WHEN 10 HYPOTHESES ARE TESTED USING EVERY ITEM OF SCALE # 1(b)-1: PEER RELATIONS (OPPOSITE SEX--MARRIAGE)

Item of Scale # 1(b)-1	Criterion Items 13	20	95	157	181	192	199	209	210	220
1.(105)*	00	00	00	00	01	27	63	00	00	00
2.(107)*	00	48	01	27	32	26	37	02	00	00
3.(111)*	00	00	00	00	00	14	00	09	00	00
4.(112)*	00	00	00	14	26	03	48	01	00	00

Note: Decimal is understood before each entry.

Underlined entries indicate non-significant F-ratios when $\alpha = .05$.

* Serial number of the item in the questionnaire.

It will be recalled that Scale # 3: Social Consciousness, unlike the remaining scales, pertains to Sense of Mission rather than that of Mutuality, the two major premises of the Youth Research Centre Survey used in this research. Therefore, doubt was expressed earlier as to whether its items will substantiate the ten predictions and the underlying rationale. The actual results from item-wise analyses are given in Table 3.29.

By examining the F-ratios in Table 3.29, it is noted that there were four or more non-significant (underlined) entries in nine of the ten columns. This demonstrates that the overall rationale and the ten hypotheses arising from it are not relevant to this scale.

Examination of the F-ratios in each row shows that 8 of the 12 items (rows) contain four or more non-significant entries, casting doubt about the items. However, if the basic rationale itself is not appropriate, an examination of the relevance of the items on the basis of substantiative predictions is superfluous.

In summary, the item-within-scale approach followed above demonstrates the usefulness of examining the external component of construct validity. In certain cases, it leads to a refinement of the scales so that each item in the refined scale confirms a number of rational predictions. Scale # 1(b)-1, one of the two scales resulting from the subdivision of a larger scale, was found to be borderline in this context,

TABLE 3.29

PROBABILITIES FOR F-RATIOS WHEN 10 HYPOTHESES ARE TESTED USING EVERY ITEM OF SCALE # 3: SOCIAL CONSCIOUSNESS

Item of Scale # 3	Criterion Items 13	20	95	157	181	192	199	209	210	220
1. (8)*	00	00	00	00	38	24	76	00	00	00
2. (10)*	70	42	00	20	98	10	10	78	12	99
3. (17)*	00	00	00	00	74	00	45	00	00	00
4. (36)*	04	00	00	04	47	24	45	00	00	06
5. (61)*	00	00	00	00	00	18	49	00	00	00
6. (63)*	72	23	00	00	40	42	00	52	33	00
7. (73)*	00	00	00	00	00	00	98	00	00	00
8. (88)*	00	29	00	53	41	00	74	14	49	06
9. (92)*	85	76	00	80	00	00	10	85	86	00
10. (95)*	14	00	00	00	00	26	00	00	00	00
11. (103)*	00	00	00	06	00	00	21	00	89	00
12. (113)*	74	82	00	00	50	00	13	18	08	08

Note: Decimal is understood before each entry.

Underlined entries indicate non-significant F-ratios when $\alpha = .05$.

* Serial number of the item in the questionnaire.

whereas Scales # 1(b)-2, # 1(d) and # 1(e) experienced modest changes and Scales # 1(a), # 1(c), # 2(a) and # 2(b) no change at all.

Scale # 3 pertained to a different underlying rationale than did the others and was employed as the basis for "divergent validation" (Cronbach, 1971) of the approach. It, therefore, was expected that if the method was useful, it would provide sufficient evidence to warrant its removal from further consideration with respect to the underlying rationale employed in this study. This, in fact, was what actually happened. The rationale was deemed to be unrelated to the scale and the individual items failed to support the ten predictions.

The above, therefore, attests to the capacity of this method to improve construct validity. In addition, it is a simple and straight forward approach which can be routinely employed by test developers.

CHAPTER IV

SUMMARY AND CONCLUSIONS

Summary

The intent of this research was to examine the contribution made when the 'substantive', 'structural' and 'external' components of construct validity (Loevinger, 1957) are employed in test development. This was done sequentially and systematically, using certain aspects of the affective domain.

The data used consisted of the responses of 7050 subjects in a nationwide survey in the U. S. A. conducted by the Youth Research Centre, Minneapolis, Minnesota. The subjects comprised an almost random sample of youth ranging in age from 15 to 18 years. The instrument used was also developed by the Youth Research Centre (Strommen and Gupta, 1970). It consisted of 420 items related to youths' perceptions about themselves, their families, friends and the world around them.

The substantive component was applied to develop eight scales. The first step in this was to define the constructs rationally, using existing literature in adolescent psychology in the area of anxiety and concerns as a guide. The concerns were related to:

1(a): Peer Relations (General)
1(b): Peer Relations (Opposite Sex)

1(c): Lack of Self-Confidence
1(d): School Problems
1(e): Personal Worth
2(a): Family Unity
2(b): Parental Understanding
3: Social Consciousness

All except the last one above reflect Mutuality: the rationale underlying interpersonal relations of the youth. The last construct is closer to what the Youth Research Centre has termed Sense of Mission.

One hundred and thirteen of the 420 items were selected by the researcher for their relevance to the above listed eight constructs. They were then given to a panel of eleven judges (graduate students) who were also familiarized with the constructs. Their task was to sort each item into one or more of the suggested categories or constructs. They were also free to reject any of the items.

The structural component of construct validity addressed itself to the question: 'Once a scale has been developed rationally and its content validity assured, can it be refined further through examining it psychometrically with respect to the nature of the intercorrelations of the items assigned by the judges to it? That is, through assessing its structure in terms of homogeneity?' This was done through using sequentially two steps:

(1) Item variance-covariance matrix of each scale was cluster analyzed, using DuBois, Gleser and Loevinger (1952) approach as programmed for IBM 360 by Gupta and Burnett (1972). It also gives KR-20. The possible results of this approach included: (a) the reduction in the size of the scale and, simultaneously, improvement in its reliability; (b) split of the scale into two or more parts, with varying levels of reliability; (c) elimination of a scale for having very low reliability.

(2) The reliability of the scales resulting from (1) above was maximized through differential response weighting, using the method of reciprocal averages (Mosier, 1943, 1946).

It was expected that the above would lead to a considerable amount of improvement in the internal consistency of each of the scales.

The external component of construct validity was used to ensure that the scales possessed high criterion-related validity. It was not possible to employ the usual correlational approaches for this purpose since measures purporting to assess the same constructs were not available, nor did we have actual samples of relevant behaviors. The alternative procedure employed here involved making rational predictions and then examining these empirically.

The rationale underlying the predictions was that of Mutuality. Criterion groups were formed on the basis of the subjects' responses to ten items giving personal data. Since Scale # 3 was related to Sense of Mission rather than to Mutuality, the predictions were not expected to be supported in its case, thereby providing evidence of divergent validity. The predictions were examined in two ways: (1) using the entire scale and (2) using each item within each scale. It was also expected that the scales would improve if the items which did not support the predictions were dropped.

Results and Conclusions

The substantive validity of the rationally described constructs was established by a high degree of agreement between the judges in assigning each of the 113 items to the eight constructs. It was arbitrarily decided that an item would be assigned to a scale if 6 of the 11 judges were in agreement on the construct. In fact, this minimum level of agreement was used for only 10 of the 113 items. Only four items remained unassigned on this ground. The judges were in complete agreement on 31 of the 113 items; on 20 other items, the judges had 91% agreement; and on 29 items, the agreement was 82%. In total, on 80 of the 113 items, the judges had 82% or higher agreement. Only one item was assigned to two scales. Such a high level of agreement was perhaps the result of this researcher having spent a great deal of time and effort in formulating the constructs and selecting the related items. The high agreement could also mean that the rationale of the constructs was clear so that the substantive validity of the scales was high.

The use of several judges rather than depending on the judgment of the investigator himself was found to improve the substantive validity of the scales.

Another important aspect of using a large number of trained judges lies in the fact that the researcher is obligated to fully develop and clearly explain his rationale. If the agreement among the judges is very low on a substantial number of items, this could be taken to mean that perhaps the rationale itself is either not clearly expounded or that the items do not adequately represent the 'universe' as defined by the rationale. In either case, the researcher must reconsider his rationale, the constructs, the universe implied by each of the constructs, and the particular items selected for them.

In view of the fact that the judges were in high agreement about the assignment of the items to the various constructs, one could say that the rationale for the constructs used here was clear and that the substantive validity of the scales was high, thus providing evidence in support of the benefits from using this particular approach as a first step in test development.

It was expected that refining the scales through examining their structure would improve the internal consistency and thereby the psychological interpretability of each of the judges' scales.

Cluster analysis did not result in any change in four scales. In their case, it served as an objective way for verifying the judges' assignments. Of the remaining four which were effected, Scales # 1(a) and # 3 were shortened by deletion of one item each which was not found highly related to the remaining items in the scales. Scale # 1(d) was reduced in length by two items for the same reason. When the content of the rejected items was examined, it appeared to be logical why they should be dropped anyway. Close examination of the two items dropped from Scale # 1(d), for example, suggested that they emphasized different aspects than those items remaining in the scale. Their removal made the interpretation of the scores on this scale more meaningful.

Scale # 1(b) was found to have 'bi-polar' structure when cluster analyzed. The two resulting scales related to two somewhat different aspects of concerns over peer relations with the opposite sex--1(b)-1: Peer Relations (Opposite Sex--Marriage) and 1(b)-2: Peer Relations (Opposite Sex--Social). It can thus be concluded that cluster analysis led to improving the reliability and the psychological interpretability of the scales.

The field worker, using tests so confirmed, could feel greater assurance that his constructs are well substantiated.

The use of differential weights for the item responses slightly improved the internal consistency on each scale. Its use can be recommended, therefore, in those situations in which computer facilities are available to the test developer.

The external component of construct validity was sought to be ensured through making rational and theoretically justifiable predictions and then empirically examining them.

The predictions were confirmed by seven of the nine scales, the exceptions being Scales # 1(b)-1 and # 3 only. It will be recalled that Scale # 1(b)-1 had resulted from a sub-division of Scale # 1(b). In the light of this evidence, its relevance to the rationale of Mutuality should be reexamined. Since Scale # 3 did not pertain to the rationale of Sense of Mission, it was not expected to support the rational predictions. This actually turned out to be the case. Thus, there is some indirect evidence in favor of its validity.

When the predictions were examined on the basis of each item within each scale, it was found that the items of Scale # 1(b)-1 were borderline, thus providing further reason to doubt the validity of this scale. Scales # 1(b)-2,

1(d) and # 1(e) experienced nominal changes by the removal of one or two items which did not confirm three or more of the ten predictions. Scales # 1(a), # 1(c), # 2(a) and # 2(b) experienced no change at all.

Just as Scale # 3 failed to confirm the predictions, so did the individual items. This can be taken to mean definite evidence about the validity of the scale.

The above, therefore, attests to the contribution of the item-wise approach to improving the 'predictive validity' of the scales through both convergent and divergent validation. The method is conceptually simple and straightforward and is well worth using.

Limitations of the Study and Suggestions for Further Research

A study of the type undertaken here requires a large amount of data. The Youth Research Centre Survey data, with its 420 items and 7050 subjects, were ideal for the purpose. Also, since the study is methodological in nature, the substantive aspect is of secondary importance, so is the nature of sampling procedures. Therefore, the fact that the subjects were non-Canadian has little bearing on the generalizability of the findings.

A limitation of this study lay in the fact that it used available data. The items (the responses to which constituted the data) determined to a large extent the rationale one could employ. This, however, does not affect the psychometric findings, but this approach should be used in other contexts also, for example, in a totally new situation in which items are not available and then see whether the judges still have high degree of agreement.

It may also be of interest to conduct a study comparing the scales obtained through cluster analysis of the entire item pool with the scales which are derived from the judges' assignments. It is anticipated that if the underlying rationale is clearly formulated and explained to the judges, then the scales formed from these judges' assignments should be the same as those derived from the cluster analysis of the entire item pool.

The item-wise analysis of the external component appears to be one of the major contributions of this study. Its utility, however, should also be examined further in the construction of tests under different circumstances when external or follow-up data are available.

REFERENCES

1. Anastasi, A. "Some Current Developments in the Measurement and Interpretation of Test Validity" in A. Anastasi (Ed.) Testing Problems in Perspective. Washington: American Council on Education, 1966, pp. 307-317.

2. American Educational Research Association. Technical Recommendations for Achievement Tests. Washington: AERA, 1955.

3. American Psychological Association. Standards for Educational and Psychological Tests and Manuals. Washington: APA, 1966.

4. American Psychological Association, American Educational Research Association, & National Council on Measurements Used in Education. "Technical Recommendations for Psychological Tests and Diagnostic Techniques," Psychological Bulletin, 1954, 51 (Suppl.).

5. Astin, A. "Criterion-Centered Research," Educational and Psychological Measurement, 1964, 24, 807-822.

6. Baker and Schulberg. "The Development of a Community Mental Health Ideology Scale," Community Mental Health Journal, 1967, 3(3), 216-225.

7. Baldwin, A. L. Theories of Child Development. New York: John Wiley and Sons, Inc., 1967.

8. Bay, K. S. "Sampling Distributions of Reliability Estimates," Unpublished Doctoral Dissertation, University of Alberta, Edmonton, 1971.

9. Bechtoldt, H. P. "Construct Validity: A Critique," American Psychologist, 1959, 14, 619-629.

10. Bloom, B. S. (Ed.) Taxonomy of Educational Objectives. Handbook I. Cognitive Domain, New York: David McKay, 1956.

11. Bloom, B. S. et al. Taxonomy of Educational Objectives. Handbook II. Affective Domain, New York: David McKay, 1964.

12. Bower, A. C. and Anderson, C. C. "A Study of the Construct Validity of the Conceptual Systems Test," *Canadian Journal of Behavioral Science*, 1970, 2, 279-292.

13. Brodbeck, M. "The Philosophy of Science and Educational Research," *Review of Educational Research*, 1957, 27 427-440.

14. Campbell, D. T. "Recommendations for APA Test Standards Regarding Construct, Trait or Discriminant Validity," *American Psychologist*, 1960, 15, 546-553.

15. Campbell, D. T. & Fiske, D. W. "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," *Psychological Bulletin*, 1959, 56, 81-105.

16. Cain, Arthur H. *Young People and Neurosis*. New York: The John Day Company, 1970.

17. Cattell, R. B. "Validity and Reliability: A Proposed More Basic Set of Concepts," *Journal of Educational Psychology*, 1964, 55, 1-22.

18. Cattell, R. B. "The Importance of Factor-Trueness and Validity versus Homogeneity and Orthogonality in Test Scales," *Educ. and Psychol. M.*, 1964, 24(1), 3-30.

19. Cattell, R. B. "Hypothesis on the Principle Personality Dimensions in Children and Tests Constructed for Them," *J. Genetic Psychology*, 1962, 101, 145-163.

20. Cattell, R. B. *Personality and Motivation Structure and Measurement*. Yonkers, N. Y.: World Book Company, 1957.

21. Coombs, C. H. "Theory and Methods of Social Measurement" in L. Festinger and D. Katz (Eds.). *Research Measurement in the Behavioral Sciences*. New York: Dryden, 1953, pp. 471-535.

22. Cronbach, L. J. & Meehl, P. E. "Construct Validity in Psychological Tests," *Psychological Bulletin*, 1955, 52, 281-302.

23. Cronbach, L. J. "Validity" in R. L. Thorndike (Ed.) *Educational Measurement*. Washington: American Council on Education, 1971, pp. 443-504.

24. Cronbach, L. J. Essentials of Psychological Testing. 3rd. Ed. New York: Harper and Row, 1970, 752pp.

25. Cronbach, L. J. "Coefficient Alpha and the Internal Structure of Tests," Psychometrika, 1951, 16, 297-334.

26. Cureton, E. E. "Validity" in E. F. Lindquist (Ed.) Educational Measurement. Washington: American Council on Education, 1951, pp. 621-694.

27. Cureton, Edward E. "Validity, Reliability and Baloney," Educational and Psychological Measurement, 1950, 10, 94-96.

28. DuBois, P. H., Gleser, G. C. and Loevinger, J. "Maximizing the Discriminating Power of a Multiple-Score Test," Psychometrika, 1953, 18, 309-317.

29. DuBois, P. H. et al. The Construction of Homogeneous Keys for a Biographical Inventory. Res. Bull. 52-18, Air Training Command, Human Resources Research Centre, 1952, Lackland Air Force Base.

30. Dunnette, M. D. "A Note on the Criterion," J. Appl. Psychol., 1963(b), 47, 251-254.

31. Ebel, R. L. "Must All Tests Be Valid?" American Psychologist, 1961, 16, 640-647.

32. Ellsworth, Rudolf Allen "Validation of Two Projective Instruments for Assessment of Children's Self-Concept," Dissert. Abst. Int., 1971(Apr.), Vol. 31, (15-A), 5196.

33. Erikson, Erik H. Identity versus Identity Diffusion in Mitchell, J. Relevant Readings in Human Development, Berkeley: McCutcheon Publishing Corporation, 1971.

34. Erikson, Erik H. Childhood and Society. New York: W. W. Norton and Company, 1950.

35. Feigl, H. "Philosophical Embarrassments of Psychology," American Psychologist, 1959, 14, 115-128.

36. Frank, P. (Ed.) The Validation of Scientific Theories. Boston: Beacon, 1956.

37. Friedenberg, Edgar Z. The Vanishing Adolescent. New York: Dell Publishing Company.

38. Garrison, Karl C. Psychology of Adolescence. Englewood Cliffs, New Jersey: Prentice-Hall Inc., 1965.

39. Gessel, A. et al. Youth--The Years from Ten to Sixteen. New York: Harper and Brothers, 1956.

40. Grice, G. R. "Dependence of Empirical Laws on the Source of Experimental Variation," Psychological Bulletin, 1966, 66, 488-498.

41. Guilford, J. P. Personality. New York: McGraw-Hill 1959.

42. Guilford, J. P. "Sixteen Divergent-Production Abilities at the Ninth Grade Level," Multivariate Behavioral Research, 1966, I(1), 43-66.

43. Guilford, J. P. "Predicting Achievement in Ninth Grade Mathematics from Measures of Intellectual-Apptitude Factors," Educ. and Psychol. Meas., 1965, 25(3).

44. Guion, Robert M. Personnel Testing. New York: McGraw-Hill, 1965.

45. Gulliksen, H. Theory of Mental Tests. New York: Wiley 1950.

46. Gulliksen, Harold O. "Intrinsic Validity," American Psychologist, 1950, 5, 511-517.

47. Gupta, R. K. "Certain Techniques of Multivariate Analysis Applied to Different Measures of Inter-Item Relationships for Developing Unifactor Tests," Journal of Educational Measurement, 1968, 5, 223-230.

48. Gupta, R. K. & Burnett, J. D. "A Program to Carry Out Cluster Analysis by Homogeneous Grouping," Educational and Psychological Measurement, 1972, 32, 185-190.

49. Harman, H. Modern Factor Analysis. Chicago: University of Chicago Press, 1967.

50. Hempel, C. G. Philosophy of the Natural Sciences. Englewood Cliffs, N. J.: Prentice-Hall, 1966.

51. Hempel, C. G. Aspects of Scientific Explanation and Other Essays. New York: Free Press, 1965.

52. Hill, A. H. "Use of a Structured Autobiography in Construct Validity of Personality Scales," J. Consult. Psychol., 1967, 31(6), 551-556.

53. Hoephner, R. "A Construct Item Analysis on Some Aptitude Tests," *Educ. and Psychol. M.*, 1967, 27(2), 289-299.

54. Hoyt, C. J. "Test Reliability Estimated by Analysis of Variance," *Psychometrika*, 1941, 6, 153-160.

55. Israel, H. M. "A Measurement of Affect Discrimination: A Validation Study," *Dist. Abst. Int.*, 1970 (Oct.), Vol. 31 (4-B), 2282-2283.

56. Jackson, Douglas "Multimethod Factor Analysis in the Evaluation of Convergent and Discriminant Validity," *Psychological Bulletin*, 1969, 72(1), 30-49.

57. Jessor, R. and Hammond, K. R. "Construct Validity and the Taylor Anxiety Scale," *Psychol. Bull.*, 1957, 54, 3, 161-170.

58. Kee, Herbert W. "The Ph.D. in Industry--or, What's a Nice Guy Like You Doing in a Place Like This," *The Canadian Psychologist*, Vol. 12, No. 4, Oct. 1971.

59. Kerlinger, F. N. *Foundations of Behavior Research.* New York: Holt, Rinehart and Winston, Inc., 1964, 739pp.

60. Kruskal, J. B. "Multi-Dimensional Scaling by Optimizing Goodness-of-Fit to a Nonmetric Hypothesis," *Psychometrika*, 1964(a), 29, 1-27.

61. Kruskal, J. B. "Nonmetric Multi-Dimensional Scaling: A Numerical Method," *Psychometrika*, 1964(b), 29, 115-129.

62. Kohout, Frank James, "The Organization of Self Conceptions: Construct Validity of the Twenty Statements Test," *Dissert. Abst. Int.*, 1971 (Jan.), Vol. 31 (7-A), 3639.

63. Kuder, G. F. and Richardson, M. W. "The Theory of the Estimation of Test Reliability," *Psychometrika*, 1937, 2, 151-160.

64. Levine, Frederic M. "The Empirical Anchor Technique: A Simple Experimental Method for Testing the Construct Validity of Factor Titles," *Multivariate Behavioral Research*, 1967, 2(2), 251-254.

65. Levitt, Eugene E. *The Psychology of Anxiety.* New York: The Bobbs-Merril Company, Inc., 1967.

66. Loevinger, J. "Objective Tests as Instruments of Psychological Theory," *Psychological Reports*, 1957, 3, 635-694.

67. Loevinger, J. "Some Principles of Personality Measurement," *Educ. and Psychol. M.*, 1955, 15, 3-17.

68. Loevinger, J. "The Technic of Homogeneous Tests Compared with Some Aspects of Scale Analysis and Factor Analysis," *Psychol. Bull.*, 1948, 45, 507-530.

69. Loevinger, J. "A Systematic Approach to the Construction and Evaluation of Tests of Ability," *Psychological Monographs*, 1947, 61, 49pp.

70. Lord, F. M. and Novick, M. R. *Statistical Theories of Mental Test Scores*. Reading Mass: Addison-Wesley Publishing Company, 1968.

71. MacCorquodale, K. and Meehl, P. E. "On a Distinction Between Hypothetical Constructs and Intervening Variables," *Psychological Review*, 1948, 55, 95-07.

72. Magnusson, David. *Test Theory*. Don Mills, Ontario: Addison-Wesley Publishing Company, 1967.

73. Margenau, H. *Nature of Physical Reality*. New York: McGraw-Hill, 1950.

74. Matteson, M. T. "An Exploratory Investigation of a Methodology for Constructing Homogeneous Keys for a Biographical Inventory," *Dissert. Abst. Int.*, 1970 (May), Vol. 30 (11-13), 5274.

75. McLeish, J. *Students Attitudes and College Environments*. Cambridge: Cambridge Institute of Education, 1970, 251pp.

76. Mead, Margaret *The Coming of Age in Samoa*. New York: New American Library, 1950.

77. Melton, A. W. "Individual Differences and Theoretical Process Variables" in R. M. Gagni (Ed.). *Learning and Individual Differences*. Cleveland: Charles E. Merril, 1966, pp. 238-252.

78. Meyer, W. J. *Childhood and Adolescence*. Toronto: Blaisdell Publishing Company, 1967.

79. Mitchell, J. *Adolescence--Some Critical Issues.* Toronto: Holt, Rinehart and Winston of Canada, Ltd., 1971.

80. Mosier, Charles I. "On the Reliability of a Weighted Composite," *Psychometrika*, 1943, 8, 161-168.

81. Mosier, Charles I. "A Critical Examination of the Concept of Face Validity," *Educational and Psychological Measurement*, 1947, I, 191-205.

82. Mosier, C. "Machine Methods of Scaling by Reciprocal Averages," *Proceedings of Research Forum*. International Business Machine Corporation, 1946.

83. Muus, Rolf E. *Theories of Adolescence*. New York: Random House, 1968.

84. Nagel, E. *The Structure of Science*. New York: Harcourt, Brace and World, 1961.

85. Nunnally, J. C. *Psychometric Theory*. New York: McGraw-Hill Book Company, 1967, 640pp.

86. Ogston, D. and Drakeford, G. "Note on the Costello-Comrey Anxiety Scale (CCAS)," *Psychological Reports*, 1971, 28, 287-289.

87. Peak, H. "Problems of Objective Observation" in L. Festinger & D. Katz (Eds.), *Research Methods in the Behavioral Sciences*. New York: Dryden Press, 1953, pp. 243-300.

88. Podrasky E. F. "The Development and Validation of a Quasi-Nonverbal Test for Measuring Attainment of Relational Concepts in High School Physics," *Dissert. Abst. Int.*, 1971 (July), Vol. 32, (1-A), 253.

89. Rozeboom. W. W. *Foundations of the Theory of Prediction*. Homewood, Illinois: The Dorsey Press, 1966, 628pp.

90. Rozeboom, W. W. *The Art of Metascience* in Royce, Joseph R. *Toward Unification in Psychology*. Toronto: University of Toronto Press, 1970.

91. Rozeboom, W. W. "Scaling Theory and the Nature of Measurement," *Synthese*, 16, 1966, 170-233.

92. Rozeboom, W. W. "Mediation Variables in Scientific Theory," *Psychological Review*, 1956, 63, 249-264.

93. Sabers, Darrell and White, Gordon W. "The Effects of Differential Weighting of Individual Item Responses on the Predictive Validity and Reliability of an Aptitude Test," *Journal of Educational Meas*. 6(2) 1969, pp. 93-96.

94. Shevel, L. R. and Whitney, D. R. "Predictive Validity of the Mathematic Placement Examination," *Educ. and Psychol. M.*, 1969, 29(4), 895-901.

95. Spiegel, D. "Factor Analysis of 78 Variables from Nine Personality Tests and Scales," *Journal of Projective Techniques*, 1969, Vol. 33(2), 160-167.

96. Stanley, J. C. "Reliability" R. L. Thorndike (Ed.) *Educational Measurement*. Washington: American Council on Education, 1971, pp. 356-443.

97. Strang, Ruth *The Adolescent Views Himself*. New York: McGraw-Hill Book Company, 1957.

98. Strommen, M. P. & Gupta, R. K. *Manual for Youth Research Survey Section 4*. Minneapolis: Youth Research Centre, 1971.

99. Suppes, P. & Zinnes, J. L. "Basic Measurement Theory," in Luce, R. R., Bush, R. R., & Galanter, E. *Handbook of Mathematical Psychology*, Vol. I, Wiley, 1963.

100. Thorndike, R. L. "Reliability" in E. F. Lindquist (Ed.) *Educational Measurement*. Washington: American Council on Education, 1951, pp. 572-620.

101. Torgerson, W. S. *Theory and Methods of Scaling*. New York: Wiley, 1958.

102. Vernon, P. E. *Personality Assessment: A Critical Survey*. New Fetter Lane: Methuen and Co., Ltd., 1964, 333pp.

103. Wang, M. W. and Stanley, J. C. "Differential Weighting: A Review of Methods and Empirical Studies," *Review of Educational Research*, 1970, 40, pp. 663-705.

104. Wattenburg, William W. *The Adolescent Years*. New York: Harcourt Brace & World Inc., 1955.

105. Webb, et al. Unobstrusive Measures: Nonreactive Research in the Social Sciences. Chicago: Rand McNally, 1966.

106. Whiley, D. E. "Latent Partition Analysis," Psychometrika, 1967, 32, 183-199.

107. Whiting, J. W., Kluckholn, R. and Anthony, A. The Function of the Male Initiation Ceremonies at Puberty, in Rebelsky, F. and Dorman, L. Child Development and Behavior. New York: Alfred A. Knopf Inc., 1970.

108. Youth Research Centre. Youth Research Survey. Y.R.C. Minneapolis: 1970.

APPENDIX A

INSTRUCTIONS FOR RATERS

INSTRUCTIONS FOR RATERS

I. DEFINITION OF "CONCERNS"

"Concerns" relate to things which either bother young people a great deal or which young people wonder about quite often. The ones of interest in this research fall into three major categories. The major categories include the concerns which an individual has relating to (1) himself, (2) his family and (3) his country. The first two can be subdivided as cited below.

1. Concerns which an individual has relating to himself.

   (a) Peer Relations--general

   As the child develops from childhood through adolescence, his need for 'belonging' shifts from the family circle as its focal point to include his peer group until finally the peer group becomes the centre of focus. While he never abandons his family and parents completely, he broadens his horizons to include more of his peers. Failure to become accepted as a part of a group often results in frustration and insecurity.

   (b) Peer Relations--opposite sex

   In the North American society, appropriate sex role plays an important part in the life from infancy to adulthood. This is reflected by the way people dress, toys they play with, the work they do and in general, nearly all aspects of life. The change for the adolescent is more intimate since it is at this age when the body of boys and girls takes on new shape. Sexuality matures and attractions develop for the opposite sex. The adolescent is concerned with impressing the opposite sex and behaves accordingly through mannerisms, hygiene and dress. The youth is aware that he will marry and raise a family sooner or later. Concern over boy-girl relations is probably the most important and most frequent worry young people have.

(c) Lack of Self-Confidence

The concerns include fears related to social relations, that is, fear of embarrassment in public, meeting strange people and in general, situations which call for a great deal of self-confidence. When the adolescent leaves the family circle, he is faced with this sort of fear and continues to experience this fear until he gains self-confidence through success. Failure in social situations will only deepen the lack of self-confidence.

(d) School Relations and Problems

Education is increasingly more important to the adolescent if he is to function effectively and harmoniously in the present social order. The reason lies in the increasing complexity of technology and move away from farms and small family businesses. This increasing complexity of the social order has brought new demands. The youth, therefore, is obliged to acquire these skills. Failure to do so has consequences of increased lack of self-confidence. Adolescents are often frustrated by inability to concentrate and to achieve good grades. It is these sorts of things that items of concerns about school will assess.

(e) Personal Worth

Includes fears related to failure in school and personal inadequacy which generally are referred to as a sense of personal worth. Fear of failure by the young person of not having lived up to his ideals and what has been expected by friends, family and society lead to a state of anxiety and concern.

2. 'Concerns' scales which relate to family.

As the young person reaches adolescence, he desires to be free from attachments of his home and parents. He wants to be free to seek his own friends, to spend his own money and to come and go as he pleases. However, along with this freedom, he must accept responsibility for his actions which he usually isn't fully prepared to do. Thus, while he is seeking freedom, he still wants the security and protection of his family.

(a) Family Unity

Reflects individual's concern over emotional climate in the home. It includes closeness of family members and consideration for one another.

(b) Parental Understanding

Concern over a lack of communication and understanding between a youth and his parents.

3. Social Consciousness.

Youth are growing up in a society where they are free to pursue their own beliefs. The mass media and extensive educational programs make the youth aware of developments in education, politics and economics. This, coupled with the idealism of youth and their search for idealistic solutions to problems, leads them to be concerned for and even to revolt against many segments of society.

II. TASK FOR RATERS

Familiarize yourself thoroughly with each of the constructs cited above (1(a), 1(b), 1(c), 1(d), 1(e), 2(a), 2(b), 3). Then read each of the items given here and assign the item to the category you feel it best belongs. If you feel that an item fits into two or more categories, assign it accordingly. If an item does not seem to fit into any of the specified categories, put it under category 4.

I am bothered by the fact that:

	1(a)	1(b)	1(c)	1(d)	1(e)	2(a)	2(b)	3	4
1. We are not close as members of a family.	___	___	___	___	___	___	___	___	___
2. My parents seem to have forgotten how it feels to be young.	___	___	___	___	___	___	___	___	___
3. I am easily carried away by my emotions.	___	___	___	___	___	___	___	___	___
4. I become discouraged rather easily.	___	___	___	___	___	___	___	___	___
5. I have little interest in school studies.									
6. Some teachers act as though a teenager knows practically nothing.	___	___	___	___	___	___	___	___	___
7. Classmates at school could be more friendly.	___	___	___	___	___	___	___	___	___
8. Our national government often seems unresponsive to the needs of people.	___	___	___	___	___	___	___	___	___
9. I do not feel I help others enough.	___	___	___	___	___	___	___	___	___
10. Pollution of our air and water threatens to destroy all human life.	___	___	___	___	___	___	___	___	___
11. We need a greater feeling of love in our family.	___	___	___	___	___	___	___	___	___
12. It is hard to discuss my problems with my mother.	___	___	___	___	___	___	___	___	___
13. I worry about little things.	___	___	___	___	___	___	___	___	___
14. I do not take my studies seriously enough.	___	___	___	___	___	___	___	___	___

		1(a)	1(b)	1(c)	1(d)	1(e)	2(a)	2(b)	3	4
15.	Some of my teachers are unfair.									
16.	There are cliques (closed groups) in my school.									
17.	Some of the state laws now being enforced are unjust.									
18.	I am too anxious to please others.									
19.	There are not enough social activities in my home.									
20.	My parents (mother or father) nag me.									
21.	I can't seem to think straight when my sexual feelings are aroused.									
22.	I cannot think of sex in marriage as sacred.									
23.	I don't do enough to help others.									
24.	I lack confidence when reciting in class.									
25.	We don't always have recreation at school that all can enjoy.									
26.	Some of my classmates are inconsiderate of my feelings.									
27.	Friends of mine, who don't believe in war, are being forced to enter the military service.									
28.	My family is not as happy as I wish it were.									

	1(a)	1(b)	1(c)	1(d)	1(e)	2(a)	2(b)	3	4
29. My parents (mother or father) try to pry into my private life.									
30. I am not a good example to a younger sister/brother.									
31. I am afraid of marriage.									
32. I am afraid of failure or humiliation.									
33. Adults tend to underestimate the abilities of young people.									
34. School requires too much of a person's time.									
35. I lack the personality and the ability to be a leader in a group.									
36. Our world may be destroyed by a nuclear war.									
37. I cannot forgive myself for things I have done.									
38. My parents (mother or father) do not like some of my friends.									
39. I allow my feelings to overbalance my values in matters of sexual behavior.									
40. I do not know what a boy (or girl) expects when on a date.									
41. I am afraid of making mistakes.									
42. I don't know how to study well.									

	1(a)	1(b)	1(c)	1(d)	1(e)	2(a)	2(b)	3	4
43. At school I am often blamed for things I did not do.									
44. Outside of my family there is no group where I feel I really belong.									
45. I cannot keep from thinking thoughts I feel I shouldn't have.									
46. I do not know what life work to enter.									
47. The members of my family are not considerate of each other.									
48. My parents (mother or father) do not understand my dating problems.									
49. I feel that I am not as smart as others my age.									
50. There are those who are smarter than I am and get better grades.									
51. Some teachers are sarcastic and critical of what I do.									
52. In a group I often act different from what I really am.									
53. I cannot live up to the standards I have set for myself.									
54. My father and mother do not get along as they should.									

	1(a)	1(b)	1(c)	1(d)	1(e)	2(a)	2(b)	3	4
55. My parents (mother or father) do not let me make my own decisions.									
56. I find it hard to defend my moral beliefs about sex.									
57. There are not enough opportunities to be with a mixed group (boys and girls) in social activities.									
58. My feelings are easily hurt.									
59. Some of my teachers are not interested in me.									
60. My interests are often different from those of others my age.									
61. Peace among nations seems impossible.									
62. I sometimes have a feeling of superiority.									
63. Revolution and violence may destroy our country soon.									
64. My father is not as interested in me as I would like.									
65. My parents (mother or father) force religion on me.									
66. I fall in love too easily.									
67. Things sometimes do not go the way I want them to.									

	1(a)	1(b)	1(c)	1(d)	1(e)	2(a)	2(b)	3	4
68. I worry about tests.	___	___	___	___	___	___	___	___	___
69. School subjects do not offer me enough challenge.	___	___	___	___	___	___	___	___	___
70. I do not understand my parents.									
71. I sometimes think of dying or being killed.	___	___	___	___	___	___	___	___	___
72. The thought of dying frightens me.	___	___	___	___	___	___	___	___	___
73. The ideals of the Constitution are far from the realities of America today.	___	___	___	___	___	___	___	___	___
74. My parents (mother or father) are too strict.	___	___	___	___	___	___	___	___	___
75. I cannot stop liking the one with whom I broke up.	___	___	___	___	___	___	___	___	___
76. My mother is not as interested in me as I would like.	___	___	___	___	___	___	___	___	___
77. I have difficulty keeping my mind on my studies.	___	___	___	___	___	___	___	___	___
78. Some of my teachers do not understand me.	___	___	___	___	___	___	___	___	___
79. I do not easily get along with others.	___	___	___	___	___	___	___	___	___
80. I talk about people behind their backs.	___	___	___	___	___	___	___	___	___
81. We do not do things together as a family.	___	___	___	___	___	___	___	___	___
82. My parents (mother or father) do not trust me.	___	___	___	___	___	___	___	___	___

	1(a)	1(b)	1(c)	1(d)	1(e)	2(a)	2(b)	3	4
83. I don't know how girls (or boys) think.									
84. I get into moods where I can't seem to cheer up.									
85. I am not satisfied with the grades I get.									
86. I feel pressure at school to do what others do.									
87. I lack ability to participate in sports.									
88. Some people want to destroy the government because they find things to criticize.									
89. I am often jealous of my friends.									
90. I often feel lonely.									
91. I do not have many friends at school.									
92. There is so much violence and crime today.									
93. Life is such a mess--sometimes I wish I could "get away from it all".									
94. I get angry at "little things".									
95. War seems to be a useless butchery.									
96. It seems that I can never do anything right.									
97. I feel restless and always wanting to be "on the go".									
98. I often feel sorry for myself.									

	1(a)	1(b)	1(c)	1(d)	1(e)	2(a)	2(b)	3	4
99. I do not know what to do when someone makes fun of others.									
100. I am sometimes so conscious of my faults that I enjoy nothing.									
101. I am unsure of myself.									
102. I daydream too much.									
103. Quite a few in my school are experimenting with drugs.									
104. I may become seriously ill or have a crippling accident.									
I wonder about:									
105. What to look for in a life partner.									
106. Why I behave as I do.									
107. Whether I will marry someone who will give me happiness.									
108. How to keep boys/girls interested in me.									
109. Whether my sexual desires are normal.									
110. Whether I have the ability to do college work.									
111. Whether I will find a life partner.									

	1(a)	1(b)	1(c)	1(d)	1(e)	2(a)	2(b)	3	4
112. Whether I can find a life partner who feels the way I do about things that are right and wrong.	___	___	___	___	___	___	___	___	___
113. Whether it is really fair for some men to be unwilling to enter the armed services.	___	___	___	___	___	___	___	___	___

APPENDIX B

TABLES 3.14-3.19 FOR EXTERNAL COMPONENT

TABLE 3.14

MEANS AND F-RATIOS FOR SCALE # 1(b)-2: PEER RELATIONS (OPPOSITE SEX--SOCIAL) FOR CONTRASTING GROUPS FORMED ON THE TEN CRITERIA

Hypothesis	Means 1	2	3	F-ratio	Probability
1. (13)*	32.54	29.20		140.50	.00
2. (20)*	31.35	28.46		175.90	.00
3. (95)*	31.08	29.33	25.95	125.76	.00
4. (157)*	33.40	28.51		232.65	.00
5. (181)*	31.16	30.35	28.32	52.17	.00
6. (192)*	30.72	31.76	28.75	78.98	.00
7. (199)*	31.46	30.23	29.19	22.87	.00
8. (209)*	36.64	27.56		790.49	.00
9. (210)*	35.91	28.04		552.32	.00
10. (220)*	35.01	27.86		489.62	.00

* Serial number of item in the Youth Research Centre Survey

TABLE 3.15

MEANS AND F-RATIOS FOR SCALE # 1(c): SELF-CONFIDENCE FOR CONTRASTING GROUPS FORMED ON THE TEN CRITERIA

Hypothesis	Means 1	 2	 3	F-ratio	Probability
1. (13)*	37.20	32.47		320.87	.00
2. (20)*	35.48	31.44		391.69	.00
3. (95)*	34.59	32.97	29.24	149.38	.00
4. (157)*	32.24	27.34		312.16	.00
5. (181)*	35.14	33.95	31.44	95.71	.00
6. (192)*	33.72	34.62	32.66	37.31	.00
7. (199)*	35.76	34.15	32.19	72.18	.00
8. (209)*	35.18	26.51		979.04	.00
9. (210)*	37.08	26.08		1639.17	.00
10. (220)*	31.25	27.90		136.46	.00

* Serial number of item in the Youth Research Centre Survey

TABLE 3.16

MEANS AND F-RATIOS FOR SCALE # 1(d): SCHOOL PROBLEMS FOR CONTRASTING GROUPS FORMED ON THE TEN CRITERIA

Hypothesis	Means 1	Means 2	Means 3	F-ratio	Probability
1. (13)*	36.59	34.50		39.76	.00
2. (20)*	36.09	33.90		76.89	.00
3. (95)*	35.85	34.61	32.03	52.40	.00
4. (157)*	40.63	39.28		33.18	.00
5. (181)*	35.86	34.88	34.39	8.78	.00
6. (192)*	35.52	36.76	33.96	51.39	.00
7. (199)*	35.98	35.33	34.42	9.77	.00
8. (209)*	40.58	39.38		23.86	.00
9. (210)*	40.54	39.43		19.47	.00
10. (220)*	40.56	39.36		25.15	.00

* Serial number of item in the Youth Research Centre Survey

TABLE 3.17

MEANS AND F-RATIOS FOR SCALE # 1(e): SELF WORTH FOR CONTRASTING GROUPS FORMED ON THE TEN CRITERIA

Hypothesis	Means 1	2	3	F-ratio	Probability
1. (13)*	37.82	33.89		233.77	.00
2. (20)*	36.18	33.22		219.30	.00
3. (95)*	36.01	34.03	30.45	177.41	.00
4. (157)*	33.13	30.19		187.10	.00
5. (181)*	35.94	35.03	33.29	50.28	.00
6. (192)*	35.02	36.05	33.86	47.40	.00
7. (199)*	36.39	35.07	33.87	34.09	.00
8. (209)*	33.25	30.34		167.41	.00
9. (210)*	32.99	30.50		117.48	.00
10. (220)*	32.89	30.63		63.75	.00

* Serial number of item in the Youth Research Centre Survey

TABLE 3.18

MEANS AND F-RATIOS FOR SCALE # 2(a): FAMILY UNITY FOR CONTRASTING GROUPS FORMED ON THE TEN CRITERIA

	Means				
Hypothesis	1	2	3	F-ratio	Probability
1. (13)*	33.41	29.32		112.23	.00
2. (20)*	32.18	28.20		117.23	.00
3. (95)*	31.56	29.24	26.15	78.04	.00
4. (157)*	32.01	28.80		187.70	.00
5. (181)*	31.75	30.62	28.40	36.21	.00
6. (192)*	33.56	33.97	27.69	210.41	.00
7. (199)*	32.27	30.76	29.04	28.19	.00
8. (209)*	32.12	28.92		170.24	.00
9. (210)*	31.77	29.13		110.78	.00
10. (220)*	30.92	29.34		43.02	.00

* Serial number of item in the Youth Research Centre Survey

TABLE 3.19

MEANS AND F-RATIOS FOR SCALE # 2(b): PARENTAL UNDERSTANDING FOR CONTRASTING GROUPS FORMED ON THE TEN CRITERIA

		Means			
Hypothesis	1	2	3	F-ratio	Probability
1. (13)*	31.87	28.26		115.80	.00
2. (20)*	30.87	27.19		201.37	.00
3. (95)*	30.41	27.85	25.59	90.71	.00
4. (157)*	35.72	32.20		251.65	.00
5. (181)*	30.47	29.29	27.45	37.53	.00
6. (192)*	35.53	35.70	24.69	1078.06	.00
7. (199)*	30.39	29.49	28.17	18.55	.00
8. (209)*	35.11	32.66		109.58	.00
9. (210)*	34.86	32.81		73.59	.00
10. (220)*	34.31	32.92		36.11	.00

* Serial Number of item in the Youth Research Centre Survey

APPENDIX C

PROBABILITIES OF F-RATIOS FROM ANALYSIS OF EACH INDIVIDUAL ITEM

TABLE 3.23

PROBABILITIES FOR F-RATIOS WHEN 10 HYPOTHESES ARE TESTED USING EVERY ITEM OF SCALE # 1(a): PEER RELATIONS (GENERAL)

Item of Scale # 1(a)	Criterion Items 13	20	95	157	181	192	199	209	210	220
1. (6)*	00	00	00	00	00	00	26	00	00	00
2. (7)*	00	00	25	00	00	00	00	00	00	00
3. (15)*	00	00	00	00	00	00	42	00	00	00
4. (16)*	00	00	00	00	00	00	00	00	00	00
5. (25)*	00	00	00	00	00	15	00	00	00	00
6. (26)*	00	00	00	00	00	00	00	00	00	00
7. (33)*	00	00	00	00	00	00	00	00	00	00
8. (43)*	00	00	00	00	00	00	00	00	00	00
9. (44)*	00	00	00	00	00	00	00	00	00	00
10. (51)*	00	00	00	00	00	00	00	00	00	00
11. (59)*	00	00	00	00	00	00	00	00	00	00
12. (60)*	00	00	00	00	00	00	00	00	00	43
13. (78)*	00	00	32	00	00	00	00	00	00	00
14. (79)*	00	00	00	00	00	00	00	00	00	00
15. (86)*	00	00	00	00	00	00	00	00	00	00
16. (89)*	00	00	00	00	00	00	00	00	00	00
17. (91)*	00	00	00	00	00	00	00	00	00	00

Note: Decimal is understood before each entry.

Underlined entries indicate non-significant F-ratios when $\alpha = .05$.

* Serial number of the item in the questionnaire.

TABLE 3.25

PROBABILITIES FOR F-RATIOS WHEN 10 HYPOTHESES ARE TESTED USING EVERY ITEM OF SCALE # 1(b)-2: PEER RELATIONS (OPPOSITE SEX--SOCIAL)

Item of Scale # 1(b)-2	Criterion Items									
	13	20	95	157	181	192	199	209	210	220
1. (21)*	00	00	00	11	00	00	00	00	00	00
2. (22)*	00	00	00	00	17	00	00	00	00	00
3. (31)*	00	00	00	00	00	00	00	00	00	00
4. (39)*	00	00	00	00	00	00	00	00	00	00
5. (40)*	00	00	00	00	00	00	00	38	26	13
6. (56)*	00	00	00	00	00	00	00	00	00	00
7. (57)*	00	00	00	00	00	00	00	00	00	00
8. (66)*	00	00	00	00	00	00	00	00	00	00
9. (75)*	36	22	00	00	27	00	00	00	48	00
10. (83)*	00	00	00	00	00	00	00	00	00	00
11. (108)*	00	00	00	00	00	00	00	00	00	00
12. (109)*	00	00	00	00	00	00	00	00	00	09

Note: Decimal is understood before each entry.

Underlined entries indicate non-significant F-ratios when $\alpha = .05$.

* Serial number of the item in the questionnaire.

TABLE 3.26

PROBABILITIES FOR F-RATIOS WHEN 10 HYPOTHESES ARE TESTED USING EVERY ITEM OF SCALE # 1(c): SELF-CONFIDENCE

Item of	Criterion Items									
Scale # 1(c)	13	20	95	157	181	192	199	209	210	220
1. (3)*	00	00	00	00	00	00	00	00	00	00
2. (13)*	00	00	00	00	00	00	00	00	00	00
3. (18)*	00	00	00	00	00	00	00	00	00	00
4. (24)*	00	00	00	00	00	00	00	00	00	00
5. (32)*	00	00	00	00	00	00	11	00	00	00
6. (35)*	00	00	00	00	09	00	00	00	00	00
7. (41)*	00	00	00	00	00	00	00	00	00	00
8. (46)*	00	00	00	00	00	00	00	00	00	00
9. (49)*	00	00	00	00	00	00	00	00	00	00
10. (52)*	00	00	00	00	00	00	00	00	00	00
11. (58)*	00	00	00	00	00	00	00	00	00	00
12. (87)*	00	00	00	00	00	00	00	00	00	00
13. (99)*	00	00	00	00	00	00	00	00	00	00
14. (100)*	00	00	00	00	00	00	00	00	00	00
15. (101)*	00	00	00	00	00	00	00	00	00	00
16. (104)*	00	00	00	00	00	00	07	00	00	00

Note: Decimal is understood before each entry.

Underlined entries indicate non-significant F-ratios when $\alpha = .05$.

* Serial number of the item in the questionnaire.

TABLE 3.27

PROBABILITIES FOR F-RATIOS WHEN 10 HYPOTHESES ARE TESTED USING EVERY ITEM OF SCALE # 1(e): SELF-WORTH

Item of Scale # 1(e)	Criterion Items 13	20	95	157	181	192	199	209	210	220
1. (4)*	00	00	00	00	00	00	10	00	00	00
2. (9)*	00	00	00	17	00	00	00	00	00	00
3. (23)*	00	00	00	23	37	00	00	00	00	00
4. (30)*	09	00	00	00	00	00	07	00	00	13
5. (37)*	00	00	00	00	00	00	00	00	00	00
6. (45)*	00	00	00	00	00	00	00	00	00	00
7. (53)*	00	00	00	00	00	00	00	00	00	00
8. (62)*	11	00	00	22	00	19	08	00	00	00
9. (67)*	00	00	00	00	00	00	00	00	00	00
10. (94)*	00	00	00	00	00	00	00	00	00	00
11. (96)*	00	00	00	00	00	00	00	00	00	00
12. (98)*	00	00	00	00	00	00	00	00	00	00
13. (102)*	00	00	00	00	00	00	00	00	00	00
14. (106)*	00	00	00	00	00	00	00	00	00	18

Note: Decimal is understood before each entry.

Underlined entries indicate non-significant F-ratios when $\alpha = .05$.

* Serial number of the item in the questionnaire.

TABLE 3.28

PROBABILITIES FOR F-RATIOS WHEN 10 HYPOTHESES ARE TESTED USING EVERY ITEM OF SCALE # 2(b): PARENTAL UNDERSTANDING

Item of Scale # 2(b)	Criterion Items									
	13	20	95	157	181	192	199	209	210	220
1. (2)*	00	00	00	00	00	00	00	00	00	00
2. (12)*	00	00	00	00	00	00	00	00	00	00
3. (20)*	00	00	00	00	00	00	00	00	00	00
4. (29)*	00	00	00	00	00	00	00	00	00	00
5. (38)*	00	00	00	00	00	00	00	00	00	16
6. (48)*	00	00	00	00	22	00	00	00	00	00
7. (55)*	00	00	00	00	00	00	00	00	00	00
8. (64)*	00	00	00	00	00	00	00	00	00	00
9. (65)*	00	00	00	00	00	00	00	00	00	00
10. (70)*	00	19	00	00	00	00	00	00	00	24
11. (74)*	00	00	00	00	00	00	00	00	00	00
12. (76)*	00	00	00	00	00	00	00	00	00	00
13. (82)*	00	00	00	00	00	00	00	00	00	00

Note: Decimal is understood before each entry.

Underlined entries indicate non-significant F-ratios when $\alpha = .05$.

* Serial number of the item in the questionnaire.